

GenCore version 5.1.4.P5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: April 27, 2003, 19:37:10 ; Search time 69 Seconds  
(Without alignments)  
549,459 Million cell updates/sec

Title: US-09-867-753-2

Perfect score: 986  
Sequence: 1 MARSLVHDVTFYCLSYGVK.....LMLANELRADPDCCYIYVD 184

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

SPTRMBL\_21:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp.fodent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriap:\*  
17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	789	80.0	148	4	O95030 homo sapien
2	260.5	26.4	137	4	O9Y4W1
3	216	21.9	288	4	O9BRO0
4	216	21.9	288	4	O9BRO4
5	212.5	21.6	227	11	O9EQM5
6	212	21.5	286	5	O87615
7	197	20.0	382	11	O88933
8	197	20.0	387	11	O92203
9	196	19.9	314	11	O54817
10	191.5	19.4	248	13	O12952
11	191.5	19.4	562	4	O96033
12	190	19.3	640	5	O870M4
13	188.5	19.1	371	5	O61282
14	188	19.0	327	11	O70238
15	187	19.0	328	13	O9W7M5
16	186	18.9	563	5	O9VTK6

17	183	18.6	408	5	O9VPI1
18	182.5	18.5	185	13	P79857
19	181	18.4	365	6	O9GMA3
20	180	18.3	240	11	O8R4I3
21	180	18.3	299	6	O8SQ03
22	180	18.3	299	11	O9WTQ9
23	180	18.3	299	11	O9JLR8
24	179.5	18.2	362	5	O21836
25	179	18.2	301	5	O46170
26	179	18.2	391	5	O878C1
27	179	18.2	826	5	O9B130
28	177.5	18.0	276	13	O98TG7
29	177	18.0	387	13	O93582
30	176.5	17.9	493	11	O9E516
31	176	17.8	214	13	O8UVJ3
32	176	17.8	370	5	O25411
33	176	17.8	479	11	O9CX16
34	176	17.8	479	13	O13081
35	176	17.8	612	5	O24477
36	176	17.8	612	5	O26441
37	176	17.8	835	4	O96H85
38	175.5	17.8	228	5	O62546
39	175.5	17.8	371	5	O46169
40	175	17.7	484	11	O9CZK7
41	174.5	17.7	350	13	O9IAL2
42	174	17.6	282	13	O73678
43	174	17.6	295	5	O96824
44	174	17.6	464	5	O9MDA9
45	173	17.5	210	11	O9QYR0

## ALIGNMENTS

RESULT 1	
O95030	PRELIMINARY: PRT: 148 AA.
AC O95030:	
DT 01-MAY-1999 (TREMBLrel. 10, Created)	
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)	
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE WUGSC:H.GS421103.1 protein (Fragment).	
GN WUGSC:H.GS421103.1.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX NCBI_TaxID=9606;	
RP [1]	
RA Leonard S., Graves T., Coffan M.;	
RT "The sequence of Homo sapiens BAC clone GSI-42113.";	
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.	
RN [2]	
RP SEQUENCE FROM N.A.	
RA Waterston R.;	
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.	
DR EMBL: AC005023; AAC78617.1; ..	
DR HSSP: P06601; IFLJL.	
DR InterPro: IPR001356; Homeobox.	
DR Pfam: PF00046; homeobox; 1.	
DR SMART: SM00389; HOX; 1.	
DR PROSITE: PS50071; HOMEBOX_2; 1.	
FT NON_TER	
FT 148	
SO SEQUENCE	
148 AA; 16143 MW; 7C81BD318E70825C CRC64;	
Query Match	80.0%; Score 789; DB 4; Length 148;
Best local Similarity	100.0%; Pred. No. 6.3e-63;
Matches 148; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 MARSLVHDVTFYCLSYGVKISPTPOLGAASSAEGHVGAGAPGLMGNMPEGVNENGM 60
DB	1 MARSLVHDVTFYCLSYGVKISPTPOLGAASSAEGHVGAGAPGLMGNMPEGVNENGM 60



CC - SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 DR EMBL: AL590526; CAC36519.1; -;  
 DR EMBL: AK058125; BAB71675.1; -;  
 DR EMBL: AF317219; AAL02160.1; -;  
 DR EMBL: BC021719; AAH21719.1; -;  
 DR HSSP: P06601; 1FJL.  
 DR InterPro: IPR001356; Homeobox.  
 DR InterPro: IPR001933; Mitoch\_carrier.  
 DR Pfam: PF00046; homeobox.1.  
 DR ProDom: PDD00010; Homeobox.1.  
 DR SMART: SM00389; HOX.1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 DR PROSITE: PS00215; MITOCH\_CARRIER; UNKNOWN.1.  
 DR DNA-binding: Homeobox; Hypothetical protein; Nuclear protein.  
 KW SEQUENCE 288 AA; 31691 MW; 280E007064515808 CRC64;  
 SQ  
 Query Match 21.9%; Score 216; DB 4; Length 288;  
 Best Local Similarity 34.9%; Pred. No. 1.7e-11;  
 Matches 58; Conservative 23; Mismatches 57; Indels 28; Gaps 7;

QY 23 PPPLGASASAGHVGQAPGLMGNMPPGVNHNENGRDGG--MIP-----E 69  
 DB 48 PPEPGCTAGGK---LKSGAGGGEKEDG-----GEEKDGAGVPGHLMGDEGTS 98  
 QY 70 GGGGNEPQQPQPPPEP---AQAMEGPPEN--MPPTRKRTFTLLQVEELSFRRHT 125  
 DB 99 GSDGNEVDSQSKPEGGQYSRPGAVGLEPNAOQPVNH--APPLDQLERLIFORE 156  
 QY 126 QYDPVTRRELAEVLGTEDEKRVWFKNRARCRRHQRELMANEL 171  
 DB 157 QPPEELRRRLARSMVTELAVQVWFENRRARRRRQRLAMRNL 202  
 RESULT 5  
 Q9EQM5 PRELIMINARY: PRT; 227 AA.  
 AC Q9EQM5;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Homeobox protein GPBOX.  
 GN 160002601R1K OR GPBOX.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Takesaki N., McIsaac R., Dean J.;  
 RT "Gbox, a novel homeobox gene preferentially expressed in female germ  
 RT cells at the onset of sexual dimorphism in mice."  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF201698; AAG36768.1; -;  
 DR MGD: MGI:1925653; 1600026001R1K.  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox.1.  
 DR ProDom: PDD00010; Homeobox.1.  
 DR SMART: SM00389; HOX.1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 DR SEQUENCE 227 AA; 25112 MW; D39CF6795AF682DC CRC64;  
 SQ

Query Match 21.6%; Score 212.5; DB 11; Length 227;  
 Best Local Similarity 35.7%; Pred. No. 2.7e-11;  
 Matches 61; Conservative 17; Mismatches 60; Indels 33; Gaps 6;  
 QY 28 GAASSAEHVGQAPGLMGNMNP-----EGGVNHNENGRDGG----- 66  
 DB 54 GGLDGEAGGAVGGEQAQAEPAISPQAEATGGESEBENEGMEGHNAGDAGSPED 113  
 QY 67 --IPEGGCGN--OEPROGPQPPPEPAQAMGPPEN--MPPTRKRTFTLLQVEELSF 120  
 DB 114 DNIGEGGGINQGPQQAALPE-----GMNPPAGNVLNQRTRRTFTHSQALDLER 168

QY 121 VERNHPOVPPVTRRELAEVLGTEDEKRVWFKNRARCRRHQRELMANEL 171  
 DB 169 LFGENRPPSLRVRRDLARWGWVDSVQEWFKMRALFRHSH-LMFCFL 218  
 RESULT 6  
 Q8T615 PRELIMINARY: PRT; 286 AA.  
 AC Q8T615;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Paired-like homeodomain transcription factor Shox.  
 GN SHOX.  
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).  
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
 OC Branchiostoma.  
 OX NCBI\_TaxID=7739;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Jackman W.R., Jr., Kimmel C.B.;  
 RT "Coincident iterated gene expression in the amphioxus neural tube."  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF465939; AAL83210.1; -;  
 KW Homeobox; DNA-binding; Nuclear protein.  
 SQ SEQUENCE 286 AA; 31551 MW; 6BF50B62D92958C1 CRC64;  
 Query Match 21.5%; Score 212; DB 5; Length 286;  
 Best Local Similarity 32.9%; Pred. No. 3.9e-11;  
 Matches 52; Conservative 24; Mismatches 48; Indels 34; Gaps 5;

QY 28 GAASSAEH--VGQAPGLMGNMPPGVNHNENGRDGMIPG--GGGQEP---RQOP 81  
 DB 46 GAEVDVGDGLCPVGSPL-----FVGGGGENSGPENKDP 82  
 QY 82 QPPEEPAQAMGPPENMPPTRKRTFTLLQVEELSFRRHTQYDPVTRRELAEVLG 141  
 DB 83 SKTGEKKDLSPNSKDGKIKORSRRTFTLLQVEELSFRTHTYPAFMRELSORLG 142  
 QY 142 VTEDKRVWFKNRARCRRHQREL-----MLANELRAD 174  
 DB 143 LSEARVQVWFQNNRACRCRQENQLQKGMILCSALGN 180

RESULT 7  
 Q88933 PRELIMINARY: PRT; 382 AA.  
 AC Q88933;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Homeobox protein SPXL.  
 GN ESX1 OR EPX.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=97398449; PubMed=9256347;  
 RA Brandford W.W., Zhao G.O., Valerius M.T., Weinstein M.,  
 RA Birkenmeier E.H., Rowe L.B., Potter S.S.;  
 RT "Spxl, a novel X-linked homeobox gene expressed during  
 RT spermatogenesis."  
 RL Mech. Dev. 65:87-98(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97392770; PubMed=9245514;  
 RX Li Y., Lemaire P., Behringer R.R.;  
 RT "Esxl, a novel x chromosome-linked homeobox gene expressed in mouse  
 RT extraembryonic tissues and male germ cells."  
 RL Dev. Biol. 188:85-95(1997).

RN [3]  
 RA SEQUENCE FROM N.A.  
 RP Branford M.W., Potter S.S.;  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RA Yan Y.-T., Yang L., Scialvolino P.J., Wang H., Chan D.C.,  
 RA Abate-Shen C., Shen M.M.;  
 RT "Epx: a novel paired-like homeobox gene expressed in the chorion and  
 RT placenta.";  
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
 CC -i- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 DR EMBL: AF085715; AAC5366.1;  
 DR EMBL: AF017735; AAD01622.1;  
 DR HSSP: P06601; IFTL.  
 DR TRANSFAC: T03478;  
 DR MGD: MGI:1096388; Esxl.  
 DR InterPro: IPR001356; Homeobox.  
 DR InterPro: IPR000047; HTH\_repressr.  
 DR Pfam: PF00046; homeobox.1.  
 DR PRINTS: PR00031; HTHREPRESSR.  
 DR PRODOM: PD000010; Homeobox.1.  
 DR SMART: SM00389; HOX.1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 DR DNA-binding: Homeobox; Nuclear protein.  
 SQ SEQUENCE 382 AA; 43540 MW; EC92301A84DA6175 CRC64;

Query Match 20.0%; Score 197; DB 11; Length 382;  
 Best Local Similarity 34.1%; Pred. No. 1.2e-09;  
 Matches 59; Conservative 18; Mismatches 72; Indels 24; Gaps 5;  
 QY 17 YQVKSIPPTQGAASSAGHYGAGPGLMGNNPBGVNHENGMNRDGG----- 65  
 AC 054817; PRELIMINARY; PRT; 314 AA.  
 DB 84 YQEBEGFEPSRGEAAAP--VAEAPQAWNGNENLGGFLESLNAOLGENDAAPVROSILMR 140  
 QY 66 -MIEPGGGGNO-----EPHQOPPP-----PEEPAQAMEGPQEPENMQPTRTKFTLLQVE 116  
 DB 141 LMQPVAAOSSPQLPRLANPLQAPQPEDEEEDEEEOGEGPOOE-PRPRRYRICFTPIOLQ 199  
 QY 117 ELESVFRHTQYPDVPTRELAENIGVTEDKYRVWFKNRKRCRRHORELMLAN 169  
 DB 200 ELEAFQVQYDPDLFARVELARLGLPEPRVQVWFONRRARWRRLRAQAFRN 252

RESULT 8  
 ID 092203 PRELIMINARY; PRT; 387 AA.  
 AC 092203;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Homeodomain protein Epx (Fragment).  
 GN Esxl OR Epx.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV;  
 RX MEDLINE=20079279; PubMed=10611245;  
 RA Yan Y.T., Stein S.M., Ding J., Shen M.M., Abate-Shen C.;  
 RT "A Novel PF/PN Motif Inhibits Nuclear Localization and DNA Binding  
 RT Activity of the Esxl Homeoprotein.";  
 RL Mol. Cell. Biol. 20:661-671(2000).  
 CC -i- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 DR EMBL: AF017734; AAD01621.1;  
 DR HSSP: P06601; IFTL.  
 DR MGD: MGI:1096388; Esxl.  
 DR InterPro: IPR001356; Homeobox.  
 DR InterPro: IPR000047; HTH\_repressr.  
 DR InterPro: IPR000047; HTH\_repressr.

DR InterPro: IPR002965; P\_rich\_extensn.  
 DR Pfam: PF00046; homeobox.1.  
 DR PRINTS: PR00031; HTHREPRESSR.  
 DR PRINTS: PR01217; PRICHEXTENSN.  
 DR PRODOM: PD000010; Homeobox.1.  
 DR SMART: SM00389; HOX.1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 DR DNA-binding: Homeobox; Nuclear protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 387 AA; 44170 MW; 0D6A6E4C36CF9C03 CRC64;

Query Match 20.0%; Score 197; DB 11; Length 387;  
 Best Local Similarity 34.1%; Pred. No. 1.2e-09;  
 Matches 59; Conservative 18; Mismatches 72; Indels 24; Gaps 5;  
 QY 17 YQVKSIPPTQGAASSAGHYGAGPGLMGNNPBGVNHENGMNRDGG----- 65  
 DB 89 YQEBEGFEPSRGEAAAP--VAEAPQAWNGNENLGGFLESLNAOLGENDAAPVROSILMR 145  
 QY 66 -MIEPGGGGNO-----EPHQOPPP-----PEEPAQAMEGPQEPENMQPTRTKFTLLQVE 116  
 DB 146 LMQPVAAOSSPQLPRLANPLQAPQPEDEEEDEEEOGEGPOOE-PRPRRYRICFTPIOLQ 204  
 QY 117 ELESVFRHTQYPDVPTRELAENIGVTEDKYRVWFKNRKRCRRHORELMLAN 169  
 DB 205 ELEAFQVQYDPDLFARVELARLGLPEPRVQVWFONRRARWRRLRAQAFRN 257

RESULT 9  
 ID 054817 PRELIMINARY; PRT; 314 AA.  
 AC 054817;  
 DT 01-JUN-1998 (TREMBlrel. 06, Created)  
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Paired-like homeodomain containing protein.  
 GN Esxl OR Etx.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SWISS;  
 RX MEDLINE=97392770; PubMed=9245514;  
 RA Li Y., Lemaire P., Behringer R.R.;  
 RT "Esxl, a novel X chromosome-linked homeobox gene expressed in mouse  
 RT extraembryonic tissues and male germ cells.";  
 RL Dev. Biol. 188:85-95(1997).  
 CC -i- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 DR EMBL: AF004211; AAB94670.1;  
 DR HSSP: P06601; IFTL.  
 DR TRANSFAC: T03474;  
 DR MGD: MGI:1096388; Esxl.  
 DR InterPro: IPR001356; Homeobox.  
 DR InterPro: IPR000047; HTH\_repressr.  
 DR Pfam: PF00046; homeobox.1.  
 DR PRINTS: PR00031; HTHREPRESSR.  
 DR PRODOM: PD000010; Homeobox.1.  
 DR SMART: SM00389; HOX.1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 DR DNA-binding: Homeobox; Nuclear protein.  
 SQ SEQUENCE 314 AA; 35969 MW; 7256F81DA4AA246 CRC64;

Query Match 19.9%; Score 196; DB 11; Length 314;  
 Best Local Similarity 34.1%; Pred. No. 1.2e-09;  
 Matches 59; Conservative 18; Mismatches 72; Indels 24; Gaps 5;  
 QY 17 YQVKSIPPTQGAASSAGHYGAGPGLMGNNPBGVNHENGMNRDGG----- 65  
 DB 16 YQEBEGFEPSRGEAAAP--VAEAPQAWNGNENLGGFLESLNAOLGENDAAPVROSILMR 72







**This Page Blank (uspto)**





[illegible]

```

Accession: I48902; I53118; S26077
R:Kern, M.J.; Arago, E.A.; Birkenmeier, E.H.; Rowe, L.B.; Potter, S.S.
Genomics 19, 334-340, 1994
A>Title: Genomic organization and chromosome localization of the murine homeobox gene
A:Reference number: I48902; MUID:94245205; PMID:7910581
A:Accession: I48902
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-217 <RES>
A:Cross-references: EMBL:U03873; NID:g460124; PIDN:AAC52139.1; PID:g460125
R:Sejersted, P., Lilly, B.; Bryson, L.J.; Wang, Y.; Sasano, D.A.; Olson, E.N.
Development 115, 1087-1101, 1992
A>Title: Moxa: a mesodermally restricted homeodomain protein that binds an essential s
A:Reference number: I53118; MUID:93083424; PMID:1360403
A:Accession: I53118
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-217 <RES>
A:Cross-references: EMD:L06502; NID:g199583; PIDN:AAA39672.1; PID:g199584
A:Residues: 1-217 <RE>
A:Molecule type: mRNA
R:Kern, M.J.; Witte, D.P.; Valerius, M.T.; Atronow, B.U.; Potter, S.S.
Nucleic Acids Res. 20, 5189-5195, 1992
A>Title: A novel murine homeobox gene isolated by a tissue specific PCR cloning strat
A:Reference number: S26076; MUID:93027261; PMID:1383943
A:Accession: S26077
A:Molecule type: mRNA
A:Residues: 1-217 <KER>
A:Cross-references: EMBL:X59726
A:Experimental source: fetal heart
C:Genetics:
A:Gene: Pmx
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: alternative splicing; DNA binding; homeobox; nucleus; transcription regul
F:95-151/Domain: homeobox homology <HOX>

Query Match          18.7%; Score 184.5; DB 2; Length 217;
Best Local Similarity 29.0%; Pred. No. 2.8e-07;
Matches    49; Conservative   27; Mismatches    54; Indels     39; Gaps      5;

OY  31 SAABGVGAGPGLMGKNMNPES-----GYNHENGANRDGMTP-----EGG 71
        | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB  2 TSSVGHVEROPALGRDSPGNLDTLQAKKNFSYSLDLEAGDMVAQAADSVSGEAG 61
OY  72 -----GGNQPRQQPPPEEPAAAMEGQPENMQPTFRRTFTLLQVELESY 121
        | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB  62 RSLSPGLTSSSDTPQDD-----NDQINSEKKKRQRNRRTTFNSSQLALERV 112
OY  122 FRHTGPVPPTFRRLAEMLGVTEKVRVMFKRKARCRHRORELILANE 170
        | | ||| | | | | | : | : | : | : | : | : | : | : | : | : |
DB  113 FERTHPFAFYEDLARRYNLTEARVQVFQFNRRKRFRRNER-AATLAANK 160

RESULT 5
S26076
homeotic protein K-2a - mouse
C:Species: Mus musculus (house mouse)
C>Date: 25-Feb-1994 #sequence_revision 01-Sep-1995 #text_change 24-Sep-1999
C:Accession: S26076
A:Accession: S26076
R:Kern, M.J.; Witte, D.P.; Valerius, M.T.; Atronow, B.U.; Potter, S.S.
Nucleic Acids Res. 20, 5189-5195, 1992
A>Title: A novel murine homeobox gene isolated by a tissue specific PCR cloning strat
A:Reference number: S26076; MUID:93027261; PMID:1383943
A:Accession: S26076
A:Molecule type: mRNA
A:Residues: 1-245 <KER>
A:Cross-references: EMBL:X59725; NTD:g51361; PIDN:CAA42410.1; PID:g51362
A:Experimental source: fetal heart
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: alternative splicing; DNA binding; homeobox; nucleus; transcription regul
F:95-151/Domain: homeobox homology <HOX>

Query Match          18.7%; Score 184.5; DB 2; Length 245;
Best Local Similarity 29.0%; Pred. No. 3.2e-07;
Matches    49; Conservative   27; Mismatches    54; Indels     39; Gaps      5;

```





```

QY      VNHGGNNRROGMY-----PEGGGNDPFRQOPQPPPEPAOAMEG-----POPEN 100
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      VSHLLDLEAGDMAAGDBGG-----EPKSLSEBGLTSGSDTPQDQN 81
      36 VSHLLDLEAGDMAAGDBGG-----EPKSLSEBGLTSGSDTPQDQN 81

QY      101 MQ-----PYRTRKFTLLQVEELSEVFRTHTQYDVPTRRELAENLGVEDKVRVA 150
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      82 DQLNSEKKRRKQRNRNTTSSSOLAENVEFRTHTPDFAVEDLARVNLTARVYA 141
      82 DQLNSEKKRRKQRNRNTTSSSOLAENVEFRTHTPDFAVEDLARVNLTARVYA 141

QY      151 FKNRARCRRHORELMANE 170
      1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      142 FQNRRAKFRNRNR-AMLASK 160
      142 FQNRRAKFRNRNR-AMLASK 160

```

RESULT 13  
JC6540  
Placenta specific-homeobox protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 15-Oct-1999  
C:Accession: JC6540  
R:Han, J.J.; Park, A.R.; Sung, D.Y.; Chun, J.Y.  
Gene 207, 159-166, 1998  
A:Title: Pax, a novel murine homeobox gene expressed in placenta.  
A:Reference number: JC6540; MUID:98172748; PMID:9511757  
A:Accession: JC6540  
A:Molecule type: mRNA  
A:Residues: 1-247 <HMAN>  
A:Cross-references: GB:AF017453  
C:Comment: This protein is involved in controlling cell fate during embryonic development.  
C:Genetics:  
A:Gene: Pax  
C:Superfamily: mouse placenta-specific homeobox protein; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; placenta; transcription regulation  
C:151-207/domain: homeobox homology <Hox>

```

RESULT 14
I48713
Phox2 homeodomain protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 24-Sep-1999
C:Accession: I48713
R:Valarone, I.; Tissier-Safr, J.P.; Hirsch, M.R.; Martinez, S.; Goriadis, C.; Brunet, J.F.
Development 119, 881-896, 1993
A:Title: The mouse homeodomain protein Phox2 regulates Ncam promoter activity in concert
A:Reference number: I48314; MUID:94244481; PMID:7910552
A:Accession: I48713
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-280 <RFS>
A:Cross-references: EMBL:X75014; NID:9402641; PIDN:CAA52923.1; PID:9402642
C:Genetics:
A:Gene: Phox2
C:Superfamily: unassigned homeobox proteins; homeobox homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:91147/Domain: homeobox homology <HOX>

```

```

Best Local Similarity 31.2%; Pred. No. 1.7e-06;
Matches 44; Conservative 22; Mismatches 37; Indels 38; Gaps 3;

QY 23 PTPLGAASSAEHVGGACPLGCMNDEGGVHNHGMNDGGMIPREGGNGEPPROQ 82
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 48 PCPALGSSNCALGALRDHPA-----PYSAVPYK-----FFPEPSGLHERRKO--- 90
QY 83 PPPEPQAAMAEQPEPMQPRTRTKFTLLLOVEELSEVPHTQYPPVPTRELAENLGV 142
Db 91 -----RIIRITFTSAQMLEKELEKVEFAETHYVDIITREELAKIDL 129
QY 143 TEDKRVVWEKKRARCRRHOR 163
    | | : | | | : | | : | | : |
Db 130 TEARVQWVFQNRRAKFRKOR 150

RESULT 15
JC6130
paired box transcription factor Pax-6 - Ribbonworm
C:Species: Lineus sanguineus (ribbonworm)
C:Date: 16-Apr-1997 #sequence_revision 05-May-1997 #text_change 15-Oct-1999
C:Accession: JC6130
A:Accession: JC6130
A:Residue type: DNA
A:Molecule type: DNA
A:Keywords: EMBL:X95594; NID:g1296835; PIDN:CAA64847.1; PID:e222109; PID:g129
A:Title: Isolation of a Pax-6 homolog from the ribbonworm Lineus sanguineus.
Proc. Natl. Acad. Sci. U.S.A. 93, 2658-2663, 1996
A:Note: The authors translated the codon GCC for residue 338 as Asp
C:Comment: This factor is a key regulator of eye morphogenesis. It plays a role in ey
C:Genetics:
A:Gene: Pax-6
A:Introns: 162/3; 236/2; 287/1; 314/3
C:Superfamily: unassigned homeobox proteins; homeobox homology; paired box homology
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation
F:47-171/Domains: paired box homology <PBH>
F:270-326/Domains: homeobox homology <HOX>

```

Search completed: April 28, 2003, 03:35:54  
Job time : 39 secs

**This Page Blank (uspto)**

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2003, 19:37:10 ; Search time 23 Seconds  
(Without alignments)  
235.383 Million cell updates/sec

Title: US-09-867-753-2

Perfect score: 986  
Sequence: 1 MARSLVHDYFYCLSYQVK.....LMLANELRADPDCCYIYVD 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/p/ptodata/1/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/p/ptodata/1/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/p/ptodata/1/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/p/ptodata/1/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/p/ptodata/1/1aa/PCITUS.COMB.pep: \*  
6: /cgn2\_6/p/ptodata/1/1aa/backfill1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	190.5	19.3	302	4	US-08-957-351-7
2	190.5	19.3	313	4	US-08-957-351-9
3	189.5	19.2	302	4	US-08-957-351-30
4	188	19.1	302	4	US-08-957-351-3
5	176	17.8	247	3	US-09-129-888-2
6	173.5	17.6	315	4	US-08-957-351-27
7	169.5	17.2	271	4	US-08-957-351-26
8	161	16.3	205	2	US-08-775-009-37
9	161	16.3	240	4	US-09-636-735A-2
10	161	16.3	240	4	US-09-636-735A-12
11	160	16.2	60	2	US-08-775-009-38
12	160	16.2	99	4	US-09-031-962D-24
13	158	16.0	436	2	US-08-958-642-4
14	158	16.0	436	3	US-08-778-394-2
15	158	16.0	436	3	US-08-778-423A-4
16	155	15.7	434	2	US-08-710-249-4
17	155	15.7	434	2	US-09-220-157A-4
18	150.5	15.3	330	2	US-08-712-948-2
19	150.5	15.3	333	2	US-08-712-948-1
20	150.5	15.3	371	2	US-08-442-809A-76
21	150	15.2	302	2	US-08-203-532F-4
22	150	15.2	302	3	US-08-950-860-16
23	150	15.2	302	4	US-09-078-465-4
24	150	15.2	302	5	PCT-US95-01882A-4
25	149	15.1	57	2	US-08-891-837B-7
26	145	14.7	303	2	US-08-203-532F-2
27	145	14.7	303	4	US-09-078-465-2

28	145	14.7	303	5	PCT-US95-01882A-2	Sequence 2, Appli
29	144.5	14.7	255	4	US-09-031-962D-4	Sequence 4, Appli
30	144	14.6	50	2	US-08-891-837B-5	Sequence 5, Appli
31	144	14.6	50	2	US-08-891-837B-6	Sequence 6, Appli
32	143.5	14.6	287	4	US-09-031-962D-2	Sequence 2, Appli
33	142.5	14.5	60	2	US-08-775-009-39	Sequence 39, Appli
34	141.5	14.4	284	2	US-08-320-148B-2	Sequence 2, Appli
35	141.5	14.4	284	3	US-08-589-028-6	Sequence 6, Appli
36	141.5	14.4	284	3	US-08-784-582-6	Sequence 6, Appli
37	141.5	14.4	284	4	US-08-785-271-6	Sequence 6, Appli
38	141.5	14.4	284	4	US-09-031-858-2	Sequence 2, Appli
39	140	14.2	147	4	US-09-605-785-336	Sequence 336, App
40	140	14.2	147	4	US-09-439-313-336	Sequence 336, App
41	140	14.2	147	4	US-09-352-616A-336	Sequence 336, App
42	140	14.2	147	4	US-09-232-149A-336	Sequence 336, App
43	138	14.0	301	4	US-09-095-117-6	Sequence 6, Appli
44	138	14.0	304	4	US-09-095-117-8	Sequence 8, Appli
45	132	13.4	66	4	US-09-031-962D-22	Sequence 22, Appli

## ALIGNMENTS

```
RESULT 1
US-08-957-351-7
Sequence 7, Application US/08957351
Patent No. 6306586
GENERAL INFORMATION:
APPLICANT: Semina, Elena
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESS: FOLEY, HONG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109-2170
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,351
FILING DATE: 24-OCT-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Arnold, Beth E.
REGISTRATION NUMBER: 35,430
REFERENCE/DOCKET NUMBER: UTA-024.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 302 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-957-351-7
Query Match 19.3%; Score 190.5; DB 4; Length 302;
Best Local Similarity 34.5%; Pred. No. 1.5e+11;
Matches 49; Conservative 23; Mismatches 45; Indels 25; Gaps 5;
OY 26 QLGAAASAEHGAGCAPGLGMMNREGVNHENGMNRDQGMTRGEGGQNOERPOORPP 85
Db 2 EFGLLSEAEAR---SPALSLIS---DAQTPHP-----QLPEHGCKGQEHSD----- 40
OY 86 EEPQAAMEGPPREN---MPPRTGRTKFTLLQVELESLSVFNNHQLQYRVPRTRELAENLG 141
```

Db 41 SEKASASLPGSPEDSLKKKORRORHTFTSOLOLEATFORNRYPDMSREIAWTN 100  
QY 142 VTEDKVWVFKNRKRCRRHOR 163  
Db 101 LTEARVWVFKNRKRCRRKR 122

## RESULT 2

US-08-957-351-9  
; Sequence 9, Application US/08957351  
; Patent No. 6306586  
; GENERAL INFORMATION:  
; APPLICANT: Semina, Elena  
; APPLICANT: Murray, Jeffrey C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/957,351  
; FILING DATE: 24-OCT-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arnold, Beth E.  
; REGISTRATION NUMBER: 35,430  
; REFERENCE/DOCKET NUMBER: UIA-024.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-7000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ. ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 313 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-957-351-9

Query Match 19.3%; Score 190.5; DB 4; Length 313;  
Best Local Similarity 34.5%; Pred. No. 1.5e-11;  
Matches 49; Conservative 23; Mismatches 45; Indels 25; Gaps 5;

QY 26 QLGAASSAEHVGAGAPGLGMNPNPEGVNHENGMNRDGMIEGGGNOEPQQOPPP 85  
Db 2 EFGLLSEAEAR---SPALSL---DAGTPHP-----QLHEGCKGGEHSD----- 40  
QY 86 EEPQAAMEGPOPN-----MOPRTRTKFTLLQVEELSVFRHTQYDPVPTRELAENLG 141  
Db 41 SEKASASLPGSPEDSLKKKORRORHTFTSOLOLEATFORNRYPDMSREIAWTN 100  
QY 142 VTEDKVWVFKNRKRCRRHOR 163  
Db 101 LTEARVWVFKNRKRCRRKR 122

## RESULT 3

US-08-957-351-30  
; Sequence 30, Application US/08957351  
; Patent No. 6306586  
; GENERAL INFORMATION:  
; APPLICANT: Semina, Elena  
; APPLICANT: Murray, Jeffrey C.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/957,351  
; FILING DATE: 24-OCT-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Arnold, Beth E.  
; REGISTRATION NUMBER: 35,430  
; REFERENCE/DOCKET NUMBER: UIA-024.01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-832-7000  
; TELEFAX: 617-832-7000  
; INFORMATION FOR SEQ. ID NO: 30:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 302 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-957-351-30

Query Match 19.2%; Score 189.5; DB 4; Length 302;  
Best Local Similarity 34.5%; Pred. No. 1.9e-11;  
Matches 49; Conservative 22; Mismatches 46; Indels 25; Gaps 5;

QY 26 QLGAASSAEHVGAGAPGLGMNPNPEGVNHENGMNRDGMIEGGGNOEPQQOPPP 85  
Db 2 EFGLLSEAEAR---TPALSL---DAGTPHP-----QLHEGCKGGEHSD----- 40  
QY 86 EEPQAAMEGPOPN-----MOPRTRTKFTLLQVEELSVFRHTQYDPVPTRELAENLG 141  
Db 41 SEKASASLPGSPEDSLKKKORRORHTFTSOLOLEATFORNRYPDMSREIAWTN 100  
QY 142 VTEDKVWVFKNRKRCRRHOR 163  
Db 101 LTEARVWVFKNRKRCRRKR 122

## RESULT 4

US-08-957-351-3  
; Sequence 3, Application US/08957351  
; Patent No. 6306586  
; GENERAL INFORMATION:  
; APPLICANT: Semina, Elena  
; APPLICANT: Murray, Jeffrey C.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP  
; STREET: One Post Office Square  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-2170  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30



```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/957,351
: FILING DATE: 24-OCT-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Arnold, Beth E.
: REGISTRATION NUMBER: 35,430
: REFERENCE/DOCKET NUMBER: UTA-024.01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-832-7000
: TELEFAX: 617-832-1000
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 302 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-957-351-3

```

Query Match	19.1%;	Score 188;	DB 4;	Length 302;
Best Local Similarity	40.6%;	Pred. No. 2.6e-11;		
Matches 41; Conservative	19;	Mismatches 31;	Indels 10;	Gaps 2

07	123	RHTQDPVYRRELALNGLVDEKVRWFKKKRRCRRHQ	163
07	82	QNRKPDMSIREELAVTNLTLEAVRWFKKKRRAKMKRRK	122
Db	28	LPEHCCKQGEHSD-----SEKASASLPGGSPEDGSLKKKQRRGRTHFTSQQLQELATF	81
07	67	IPGGGGQNGEERQOPRPPPPADAAAGGPPEN-----MPRTKRTFTLLQVELLSF	122
Db	28	LPEHCCKQGEHSD-----SEKASASLPGGSPEDGSLKKKQRRGRTHFTSQQLQELATF	81

RESULT 5  
US-09-129-888-2  
; Sequence 2, Application US/09129888B

```

: APPLICANT: CHUN, Jong Yoon
: APPLICANT: HAN, Yun Jeong
: TITLE OF INVENTION: Placenta trophoblast-specific gene
: FILE REFERENCE: 1942/29
: CURRENT APPLICATION NUMBER: US/09/129,888B
: CURRENT FILING DATE: 1998-08-06
: NUMBER OF SEQ. ID NOS.: 6
: SOFTWARE: Wordperfect 6.1 Windows
: SEQ. ID NO. 2
: LENGTH: 247
: TYPE: PRT
: ORGANISM: mouse
: US-09-129-888-2

```

Query Match	17.88;	Score 176;	DB 3;	Length 247;
Post-Translational	34.38;	Score 320;	DB 10;	Length 247;

Matches	47;	Conservative	23;	Mismatches	41;	Indels	26;	Gaps	5;
---------	-----	--------------	-----	------------	-----	--------	-----	------	----

```

0Y      34 EG-HYGGCGAGCMLGMNNPEGYNHENGNNRDGMIPEGGGG-NOEPRQRPQPPPEBQA 91
      || || ||      : : ||| :||| :
Db      99 EGRHAGDGA-----SSSEDSILEEGGQNIDQPPQDEASPD----- 136

```

```

Oy  92 AMESPOPEN--MOPRTRTKFTLLQVEELESVRHTQYDPVPTRELAENIGVTEDEKVRV 149
    :: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 137 SIRNHVNLNLQALRRTTRFTHSQDLHLEERLEQETRTYSLRARRDLARMMGVDECDYON 196

```

```
QY 150 WEKNKRACRRHQRELM 166
    ||: || :|:| ||
Db 197 WFRMRALFQRNRRLVM 213
```

RESULT 6  
US-08-957-351-27  
; Sequence 27, Application US/08957351  
; Patent No. 6306586  
; GENERAL INFORMATION:

```

1  APPLICANT:  Semina, Elena
2  APPLICANT:  Murray, Jeffrey C.
3  TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR THE
4  TITLE OF INVENTION:  DIAGNOSIS AND TREATMENT OF CATARACTS
5  NUMBER OF SEQUENCES:  33
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE:  FOLEY, HONG & ELIOT LLP
8  STREET:  One Post Office Square
9  CITY:  Boston
10 STATE:  MA
11 COUNTRY:  USA
12 ZIP:  02109-2170
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE:  Floppy disk
15 COMPUTER:  IBM PC compatible
16 OPERATING SYSTEM:  PC-DOS/MS-DOS
17 SOFTWARE:  PatentIn Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER:  US/08/957,351
20 FILING DATE:  24-OCT-1997
21 CLASSIFICATION:  435
22 ATTORNEY/AGENT INFORMATION:
23 NAME:  Arnold, Beth E.
24 REGISTRATION NUMBER:  35,430
25 REFERENCE/DOCKET NUMBER:  UTA-024.01
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE:  617-832-1000
28 TELEFAX:  617-832-7000
29 INFORMATION FOR SEQ ID NO.:  27:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH:  315 amino acids
32 TYPE:  amino acid
33 STRANDEDNESS:
34 TOPOLOGY:  linear
35 MOLECULE TYPE:  protein
36
37 US-08-957-351-27

```

Query Match	17.68;	Score 173.5;	DB 4;	Length 315;
Best Local Similarity	37.18;	Pred. No. 8.2e-10;		
Matches 39;	Conservative 17;	Mismatches 30;	Indels 19;	Gaps 2;

[illegible]

RESULT 7  
US-08-957-351-26  
: Sequence 26, Application US/08957351

GENERAL INFORMATION:

APPLICANT: Murray, Jeffrey C.

TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF CATARACTS.

CORRESPONDENCE ADDRESS:

STREET: One Post Office Square

STATE: MA  
COUNTRY: USA

ZIP: 02109-2170

MEDIUM TYPE: Floppy

```

;
OPERATING SYSTEM: PC-DOS/MS-DOS
COMMNAME: Database Release #1.0 Version #1.20

```

```

; CURRENT APPLICATION DATA:

```



LOCATION: (127) ..(136)  
NAME/KEY: HELIX 2 ..(154)  
LOCATION: (144) ..(154)  
NAME/KEY: HELIX 3  
LOCATION: (158) ..(173)  
US-09-636-735A-12

Query Match 16.3%; Score 161; DB 4; Length 240;  
Best Local Similarity 27.3%; Pred. No. 1.1e-08;  
Matches 45; Conservative 30; Mismatches 52; Indels 38; Gaps 7;

OY 15 SVYQVKISPT-----POGAASAGHV-----GQAPG--LIMGNMPE-----GGVN 55  
DB 29 AAYPLGSLPTTASPNL-SYSRPYGHLSTPYTPRANPGDSYLSQQPALSLGCPAE 87  
OY 56 HENGMDNGMIPEGGGNOEPPOPPPEPAQAAMEGPOPENMOPRTTRKFTLLQV 115  
DB 88 HPELEAD-----SEKPRLSPEPSERRPOAPAKKLKRP-----RTIYSSLOL 129  
OY 116 EELESVFRHTQYDPVPTRELAENLGVTEDKRVWFKNRRCRR 160  
DB 130 QHLNORFOHTOYALPERAQLAOLGTOVKIWFONKRSKYK 174

RESULT 11  
US-08-775-009-38  
Sequence 38, Application US/08775009  
Patent No. 5935783

GENERAL INFORMATION:  
APPLICANT: Gong, Wellong  
APPLICANT: Emanuel, Beverly S.  
APPLICANT: Budarf, Marcia L.  
APPLICANT: Roe, Bruce  
TITLE OF INVENTION: No. 5935783el Genes Mapping in the Digeorge and  
TITLE OF INVENTION: Velocardiofacial Syndrome Minimal Critical Region  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &  
STREET: One Liberty Place - 46th Floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: U.S.A.  
ZIP: 19103

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/775,009  
FILING DATE: 27-DEC-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Trujillo, Doreen Yanko  
REGISTRATION NUMBER: 35,719  
REFERENCE/DOCKET NUMBER: CH-0681  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 60 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-775-009-38

Query Match 16.2%; Score 160; DB 2; Length 60;  
Best Local Similarity 52.5%; Pred. No. 2.2e-09;  
Matches 31; Conservative 12; Mismatches 16; Indels 0; Gaps 0;

OY 104 RTRTKFTLLQVEELESVFRHTQYDPVPTRELAENLGVTEDKRVWFKNRRCRRHQ 162  
DB 2 RHRRTFTDEQLEALENLFOETKYVDVGTREOLARKVHLREKVEVWFKNRAKWRBK 60

RESULT 12  
US-09-031-962D-24  
Sequence 24, Application US/09031962D  
Patent No. 6350867

GENERAL INFORMATION:  
APPLICANT: Thomas C. Hart  
APPLICANT: Jennifer A. Price  
TITLE OF INVENTION: Methods and Compositions for Enhancing  
FILE REFERENCE: WF098-18  
CURRENT APPLICATION NUMBER: US/09/031,962D  
CURRENT FILING DATE: 1998-02-27  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 24  
LENGTH: 99  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-031-962D-24

Query Match 16.2%; Score 160; DB 4; Length 99;  
Best Local Similarity 36.9%; Pred. No. 4.2e-09;  
Matches 31; Conservative 20; Mismatches 25; Indels 8; Gaps 1;

OY 77 PROQPOPEPEPAQAAMEGPOENMOPRTTRKFTLLQVEELESVFRHTQYDPVPTREL 136  
DB 1 PRLSPESERRPOAANKLRP-----RTIYSSLOLHNLNORFOHTOYALPERAQL 52  
OY 137 AENLGVTEDKRVWFKNRRCRR 160  
DB 53 AAGLGTOVKIWFONKRSKYK 76

RESULT 13  
US-08-958-642-4  
Sequence 4, Application US/08958642  
Patent No. 5948623

GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: NOVEL METHOD FOR TESTING THE  
TITLE OF INVENTION: DIFFERENTIATION STATUS IN PANCREATIC CELLS OF A MAMMAL  
NUMBER OF SEQUENCES: 16  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/958,642  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/778,423  
FILING DATE: December 31, 1996  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 436 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-958-642-4

Query Match 16.0%; Score 158; DB 2; Length 436;  
Best Local Similarity 30.8%; Pred. No. 4.7e-08;  
Matches 41; Conservative 18; Mismatches 48; Indels 26; Gaps 3;

OY 40 GAPOLMNMNPEGVNNHNGNRRGMIPEGGGNOEPPOPPPEPAQAAMEGPOPE 99

```
Db 184 GQPTDCCQOQEGGENTNISSNG-----EDSDEAQMRLQKR 222
QY 100 NMQPRTRRTKFTLLQVLEESVFRHIOYDPVPTRELAEVLGYEDKVRVWFKNKRAKRCR 159
Db 223 KIQ--RNRTSFTQEIETALKERFERTHYPDVFARERLAAKIDLEARIQVWFSNRRAKWR 280
QY 160 RHQRELMLANELR 172
Db 281 REEK---LRNQRR 290

RESULT 14
US-08-778-394-2
; Sequence 2, Application US/08778394
; Patent No. 6028184
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NOVEL METHOD FOR TESTING THE
; TITLE OF INVENTION: DIFFERENTIATION STATUS IN PANCREATIC CELLS OF A MAMMAL
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/778,394
; FILING DATE:
; CLASSIFICATION: 514
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-778-394-2

Query Match 16.0%; Score 158; DB 3; Length 436;
Best Local Similarity 30.8%; Pred. No. 4.7e-08;
Matches 41; Conservative 18; Mismatches 48; Indels 26; Gaps 3;

QY 40 GAGGLMGNNMPEGVNHENGMNRDGMIPGGGNGDEPRROQPPPEEPAQAAMEGFPQE 99
Db 184 GQPTDCCQOQEGGENTNISSNG-----EDSDEAQMRLQKR 222
QY 100 NMQPRTRRTKFTLLQVLEESVFRHIOYDPVPTRELAEVLGYEDKVRVWFKNKRAKRCR 159
Db 223 KIQ--RNRTSFTQEIETALKERFERTHYPDVFARERLAAKIDLEARIQVWFSNRRAKWR 280
QY 160 RHQRELMLANELR 172
Db 281 REEK---LRNQRR 290

RESULT 15
US-08-778-423A-4
; Sequence 4, Application US/08778423A
; Patent No. 6071697
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: NOVEL METHOD FOR TESTING THE
; TITLE OF INVENTION: DIFFERENTIATION STATUS IN PANCREATIC CELLS OF A MAMMAL
; NUMBER OF SEQUENCES: 16
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/778,423A
```

```
; FILING DATE: December 31, 1996
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 436 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-778-423A-4

Query Match 16.0%; Score 158; DB 3; Length 436;
Best Local Similarity 30.8%; Pred. No. 4.7e-08;
Matches 41; Conservative 18; Mismatches 48; Indels 26; Gaps 3;

QY 40 GAGGLMGNNMPEGVNHENGMNRDGMIPGGGNGDEPRROQPPPEEPAQAAMEGFPQE 99
Db 184 GQPTDCCQOQEGGENTNISSNG-----EDSDEAQMRLQKR 222
QY 100 NMQPRTRRTKFTLLQVLEESVFRHIOYDPVPTRELAEVLGYEDKVRVWFKNKRAKRCR 159
Db 223 KIQ--RNRTSFTQEIETALKERFERTHYPDVFARERLAAKIDLEARIQVWFSNRRAKWR 280
QY 160 RHQRELMLANELR 172
Db 281 REEK---LRNQRR 290

Search completed: April 28, 2003, 04:05:27
Job time : 33 secs
```

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003. Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 19:24:23 ; Search time 21 Seconds

(without alignments)  
363.412 Million cell updates/sec

Title: US-09-867-753-2

Perfect score: 986

Sequence: 1 MARSLVHDTVEYCLSYGVK.....IMLANELRADPDCCYIIVD 184

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	206	20.9	331	1 SHX2_HUMAN	O60902 homo sapien
2	202.5	20.5	317	1 RX2_CHICK	O9PVX0 gallus gall
3	202.5	20.5	331	1 SHX2_MOUSE	P70390 mus musculu
4	191.5	19.4	371	1 OTX_STRPU	O26417 stronglyloce
5	190.5	19.3	302	1 PIX3_HUMAN	O75364 homo sapien
6	189.5	19.2	399	1 ALX4_MOUSE	O35137 mus musculu
7	188	19.1	302	1 PIX3_MOUSE	O35160 mus musculu
8	187	19.0	302	1 PIX3_RAT	P81062 rattus norv
9	187	19.0	299	1 CRX_BOVIN	O9XSK0 bos taurus
10	185.5	18.8	252	1 GSC_HUMAN	P56915 homo sapien
11	185.5	18.8	256	1 GSC_MOUSE	O02591 mus musculu
12	184.5	18.7	245	1 PMX1_HUMAN	P54821 mus sapien
13	184.5	18.6	245	1 PMX1_MOUSE	P43271 mus musculu
14	183	18.6	408	1 AL_DROME	O06453 drosophila
15	183	18.5	411	1 ALX4_HUMAN	O9H161 homo sapien
16	182.5	18.6	364	1 OTX_PARLI	O91611 paraentrot
17	182	18.5	237	1 SHX2_RAT	O35750 rattus norv
18	182	18.5	240	1 DLX4_MOUSE	P70436 mus musculu
19	182	18.5	346	1 RX_HUMAN	O9Y2V3 homo sapien
20	181.5	18.4	325	1 OTX_MOUSE	O09113 mus musculu
21	181	18.4	419	1 GSC_DROME	P53566 drosophila
22	180	18.3	299	1 CRX_HUMAN	O43186 homo sapien
23	180	18.3	299	1 CRX_MOUSE	O54786 mus sapien
24	179.5	18.2	317	1 PIX2_HUMAN	O95697 homo sapien
25	179	18.2	342	1 RX_DROME	O911T7 rattus norv
26	179	18.2	873	1 RX_MOUSE	O9Y2Q1 drosophila
27	178.5	18.1	317	1 PIX2_MOUSE	O61658 mus musculu
28	178	18.1	185	1 HEX1_MOUSE	P53544 brachydantio
29	178	18.1	240	1 GSC_BRARE	O35085 mus musculu
30	177	18.0	353	1 ARX_MOUSE	O95811 homo sapien
31	176.5	17.9	290	1 PMX2_HUMAN	P47239 mus musculu
32	176.5	17.9	290	1 PAX7_MOUSE	O15266 homo sapien
33	176.5	17.9	292	1 SHOX_HUMAN	

34	176.5	17.9	513	1 PIX1_DROME	O18400 drosophila
35	176	17.8	245	1 PMX1_CHICK	O05437 gallus gall
36	176	17.8	280	1 PMX1_MOUSE	O62066 mus musculu
37	176	17.8	281	1 PMX1_RAT	O62782 rattus norv
38	176	17.8	479	1 PAX3_HUMAN	P23760 homo sapien
39	176	17.8	479	1 PAX3_MOUSE	P24610 mus musculu
40	175	17.7	284	1 PMX1_HUMAN	O14813 homo sapien
41	174.5	17.7	245	1 GSC_CHICK	P53545 gallus gall
42	173.5	17.6	315	1 PIX1_MOUSE	P70314 mus musculu
43	173.5	17.6	333	1 PIX2_CHICK	O93385 gallus gall
44	173	17.5	247	1 PMX2_MOUSE	O06348 mus musculu
45	173	17.5	292	1 PIX3_XENLA	O918K3 xenopus lae

#### ALIGNMENTS

RESULT 1  
ID SHX2\_HUMAN STANDARD: PRT: 331 AA.  
AC O60902; O60903; O60465; O60467;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Short stature homeobox protein 2 (Paired-related homeobox protein SHOX) (Homeobox protein Ogl2X).  
GN SHOX2 OR SHOX OR OGL2X.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RC TISSUE=Fibroblast;  
RX MEDLINE=98151525; PubMed=9482898;  
RA Blaschke R.J., Monaghan A.P., Schiller S., Schechinger B., Rao E., Padilla-Nash H., Ried T., Rappold G.A.;  
RT "SHOX", a SHOX-related homeobox gene, is implicated in craniofacial, brain, heart, and limb development.;  
RT Proc. Natl. Acad. Sci. U.S.A. 95:2406-2411(1998).  
RL [2]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Muscle;  
RX Strussberg R.;  
RL Submitted (May-2001) to the EMBL/Genbank/DBJ databases.  
[3]  
RN SEQUENCE OF 116-331 FROM N.A. (ISOFORM 2).  
RP TISSUE=Craniofacial;  
RX MEDLINE=98133920; PubMed=9466998;  
RA Semina E.V., Reiter R.S., Murray J.C.;  
RT "A new human homeobox gene OGL2X is a member of the most conserved homeobox gene family and is expressed during heart development in mouse.";  
RL Hum. Mol. Genet. 7:415-422(1998).  
CC -!- FUNCTION: May be a growth regulator and have a role in specifying neural systems involved in processing somatosensory information, as well as in face and body structure formation.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1/SHOX2A/SHOX (shown here) and 2/SHOX2B/SHOX2; are produced by alternative splicing.  
CC -!- TISSUE SPECIFICITY: Expressed in heart, skeletal muscle, liver, lung, bone marrow fibroblast, pancreas and placenta.  
CC -!- DEVELOPMENTAL STAGE: Expressed during craniofacial development as well as in heart.  
CC -!- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.  
CC -!- SIMILARITY: SUBFAMILY.  
CC -!- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 OAR DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial  
entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
CC EMBL; AJ002367; CAA05341.1; ALT\_INIT.  
CC EMBL; AJ002368; CAA05342.1; ALT\_INIT.  
CC EMBL; BC008829; AAC08829.1; ALT\_INIT.  
CC EMBL; AF022654; AAC39662.1; ALT\_INIT.  
CC EMBL; AF023203; AAC39663.1; ALT\_INIT.  
CC HSSP; P06601; 1FJL.  
CC TRANSFAC; T04223; -.  
CC TRANSFAC; T04224; -.  
CC GeneX; HGNC:10854; SHOX2.  
CC MIM; 602304; -.  
CC InterPro; IPR000047; HTH\_repressr.  
CC InterPro; IPR003654; Homeo\_OAR.  
CC InterPro; IPR001356; Homeobox.  
CC Pfam; PF00046; homeobox; 1.  
CC PRINTS; PR00024; HOMEBOX.  
CC PRINTS; PR00031; HTHREPRESSR.  
CC PRODOM; PD000010; Homeobox; 1.  
CC SMART; SM00389; HOX; 1.  
CC PROSITE; PS00027; HOMEBOX\_1; 1.  
CC PROSITE; PS50071; HOMEBOX\_2; 1.  
CC PROSITE; PS50803; OAR; 1.  
CC Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
KW Alternative splicing.  
FT DNA\_BIND 140 199 HOMEBOX.  
FT DOMAIN 133 326 OAR.  
FT DOMAIN 60 86 POLY-GLY.  
FT VARSPIC 235 246 MISSING (IN ISOPFORM SHOX2B).  
FT CONFLICT 115 115 E -> EERRKPTKAEVATLLPEAFRFL (IN REF.  
FT CONFLICT 125 125 2).  
FT CONFLICT 244 244 E -> D (IN REF. 1; CAA05341).  
FT CONFLICT 312 312 P -> S (IN REF. 2).  
FT CONFLICT 325 325 D -> N (IN REF. 2 AND 3).  
FT CONFLICT 325 325 H -> L (IN REF. 3; AAC39663).  
SQ SEQUENCE 331 AA; 34964 MW; 55431B073B3B2250 CRC64;

Query Match 20.9%; Score 206; DB 1; Length 331;  
Best Local Similarity 34.1%; Pred. No. 3e-08;  
Matches 56; Conservative 22; Mismatches 50; Indels 36; Gaps 6;

QY 24 TPOGAASSAGHGOGAPGLMGNPNPEGVNHGNNRDMGMPDEGGGNDP----- 77  
DB 53 SPAYRAGGGGGGGGGGGGGG-----GGVG-----GGAGGGGGGSRPVRELD 99  
QY 78 -----RQGPQP-----PE-----EPAQAMEGSPQENMOPRTTRFTLLQVEELESV 121  
DB 100 GAERSRPGSPRLTEVSPELKDRKEDAK-GMEDEGOTKIKORSRTFTLEQLNELERL 158  
QY 122 FRHTQYDPVPTRELAEMLGVTEDEKVRVWFKNKRACRRHOREL 165  
DB 159 FDEHYHPDAFMRBELSRLGLSEARVQWVFQNRRAKCRKQENQL 202

RESULT 2  
RX2\_CHICK  
ID RX2\_CHICK STANDARD; PRT; 317 AA.  
AC G9PVX0;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Retinal homeobox protein Rx2 (chrax2).  
GN RX2 OR RX2.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9345958; Pubmed=10415362;

RA Ohuchi H., Tomonari S., Itoh H., Mikawa T., Noji S.;  
RT "Identification of chick rx/rx genes with overlapping patterns of  
expression during early eye and brain development.";  
RL Mech. Dev. 85:193-195(1999).  
CC -1- FUNCTION: Plays a critical role in eye formation by regulating the  
CC initial specification of retinal cells and/or their subsequent  
CC proliferation (By similarity).  
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -1- DEVELOPMENTAL STAGE: Expressed at stage 4 in the ectoderm, at  
CC stage 6 in the anterior most neural plate, at stage 7 and 8 in the  
CC anterior neural fold and at stage 9-10 in the evaginating optic  
CC vesicles. At stage 14, highly expressed in developing retina and  
CC in Infundibulum region.  
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.  
CC "BICOID" SUBFAMILY.  
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entitles requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
CC EMBL; AB020318; BAA64749.1; -.  
CC HSSP; P06601; 1FJL.  
CC InterPro; IPR003654; Homeo\_OAR.  
CC InterPro; IPR001356; Homeobox.  
CC Pfam; PF00046; homeobox; 1.  
CC PRODOM; PD000010; Homeobox; 1.  
CC SMART; SM00389; HOX; 1.  
CC PROSITE; PS00027; HOMEBOX\_1; 1.  
CC PROSITE; PS50071; HOMEBOX\_2; 1.  
CC PROSITE; PS50803; OAR; 1.  
CC Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
KW Transcription regulation.  
FT DOMAIN 33 40 OCTAPEPTIDE MOTIF.  
FT DNA\_BIND 122 181 HOMEBOX.  
FT DOMAIN 294 307 OAR.  
FT DOMAIN 300 304 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
SQ SEQUENCE 317 AA; 34056 MW; 10DDF1BEC9A24910 CRC64;

Query Match 20.5%; Score 202.5; DB 1; Length 317;  
Best Local Similarity 29.3%; Pred. No. 5.1e-08;  
Matches 54; Conservative 28; Mismatches 61; Indels 41; Gaps 5;

QY 33 AEGHVGAGAPGLM-----GNMPEGGVNHGNNRDMGMPDEGGGNDP----- 76  
DB 11 AEGAFSLSPAPARSFGNPSRLHSIEAILGFTKDGILGPPDGGASAREADKRGPR 70  
QY 77 ---PROQPPPEE-----PAQAMEGSPQENMOP-----RTTRKET 111  
DB 71 HCLPKRPAEPPEPAHEQGFQEPYCPGASPELPAGDGDGKPSDEQPKKKRRRTTFT 130  
QY 112 LLQVEELESVERHTQYDPVPTRELAEMLGVTEDEKVRVWFKNKRACRRHOREL 171  
DB 131 TYQLHELERAFEKSHYPVYSREELAMKVNLPYEVAVQWVFQNRRAKMRROKLEVSMTL 190  
QY 172 RADP 175  
DB 191 QDSP 194

RESULT 3  
SHX2\_MOUSE  
ID SHX2\_MOUSE STANDARD; PRT; 331 AA.  
AC P70390; P70369;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Short stature Homeobox protein 2 (Homeobox protein Ogl2x) (OG-12)

DE (Paired family homeodomain protein Prrx3).  
 GN SHOX2 OR OGI2X OR PRX3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC TISSUE=Embryo;  
 RA MEDLINE=98058757; PubMed=9371788;  
 RA van Schick H.S.A., Smidt M.P., Rovescalli A.C., Luitjen M.,  
 RA van der Kleij A.A.M., Asch S., Kozak C.A., Nirenberg M.W.,  
 RA Burbach J.P.H.;  
 RT "Homeobox gene Prrx3 expression in rodent brain and extraneural  
 RT tissues.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:12993-12998(1997).  
 RN [2]  
 RP SEQUENCE OF 116-331 AND 235-331 FROM N.A. (ISOFORMS 1 AND 2).  
 RC STRAIN=BA1B/C; TISSUE=Liver, and Embryo;  
 RX MEDLINE=97008065; PubMed=8855241;  
 RX Rovescalli A.C., Asch S., Nirenberg M.W.;  
 RT "Cloning and characterization of four murine homeobox genes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:10691-10696(1996).  
 RN [3]  
 RP DEVELOPMENTAL EXPRESSION.  
 RX MEDLINE=98133920; PubMed=9466998;  
 RA Semina E.V., Reiter R.S., Murray J.C.;  
 RT "A new human homeobox gene OGI2X is a member of the most conserved  
 RT homeobox gene family and is expressed during heart development in  
 RT mouse.";  
 RL Hum. Mol. Genet. 7:415-422(1998).  
 CC  
 CC -1- FUNCTION: May be a growth regulator and have a role in specifying  
 CC neural systems involved in processing somatosensory information,  
 CC as well as in face and body structure formation. May also have a  
 CC role in heart development.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1/OGI2A/PRX3A (shown here) and  
 CC 2/OGI2B/PRX3B; are produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Highly expressed in striated muscle followed  
 CC by liver, kidney, testis, brain, heart, lung and spleen.  
 CC -1- DEVELOPMENTAL STAGE: Expressed from E9 to E16 day in the heart,  
 CC otic region, maxillary and mandibular components of the first  
 CC branchial arch, nasal processes, eyelid, midbrain, medulla  
 CC oblongata, limbs, dorsal root ganglia and genital tubercle. Also  
 CC expressed in non-neuronal structures around the oral cavity and in  
 CC hip and shoulder regions and in mesenchyme surrounding the  
 CC vertebrae.  
 CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.  
 CC "BICOID" SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U66918; AAC52833.1; -;  
 CC EMBL: U67055; AAC52834.1; -;  
 CC EMBL: U65071; -; NOT\_ANNOTATED\_CDS.  
 CC EMBL: U65072; AAC52831.1; ALT\_INIT.  
 CC EMBL: U65072; AAC52832.1; ALT\_INIT.  
 CC HSSP: P06601; 1FJL.  
 CC TRANSFAC: T03310; -;  
 CC TRANSFAC: T04233; -;  
 CC MGD: MGI:1201673; Shox2.  
 CC InterPro: IPR0013654; Homeo\_OAR.  
 CC InterPro: IPR001356; Homeobox.  
 CC Pfam: PF00046; homeobox.1.  
 CC PRINTS: PR00024; HOMEBOX.

DR ProDom: PD000010; Homeobox.1.  
 DR SMART: SM00389; HOX.1.  
 DR PROSITE: PS00027; HOMEBOX\_1.1.  
 DR PROSITE: PS50071; HOMEBOX\_2.1.  
 DR PROSITE: PS50803; OAR.1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Alternative splicing.  
 FT DNA\_BIND 140 199 HOMEBOX.  
 FT DOMAIN 313 326 OAR DOMAIN.  
 FT DOMAIN 59 82 POLY-GLY.  
 FT VARSPPLIC 235 246 MISSING (IN ISOFORM 2).  
 SO SEQUENCE 331 AA; 34905 MW; D24D29YEIDP3JA025 CRC64;  
 Query Match 20.5%; Score 202.5; DB 1; Length 331;  
 Best Local Similarity 32.5%; Pred. No. 5.3e-08;  
 Matches 53; Conservative 22; Mismatches 55; Indels 33; Gaps 5;  
 QY 24 TPOLGAASAGHVGOGAPGLMGNMNPEGVHENGNNRDGMIPEGGGNOEP-----77  
 Db 52 SPAYRAAGGGGGAGGGGGG-----GG-----GGGAGGGGAGGAGGRSFRELD 99  
 QY 78 -----RQOPPP-----PE-----EPAQAMGPPENMOPRTRTKFTLLQVELESVF 122  
 Db 100 GAARSRPSPSPRLTEVSPELKDRDKDAKGMEDEGOTKIKORRSRTNFTLEQLNELERLF 159  
 QY 123 RHQYPPVPTRELAEVLGVTEDEKVRVYFKKKRRCRRHREL 165  
 Db 160 DEHYPPAFMRELSORLIGLSEARVQWFOFRRAKCRKROENQL 202  
 RESULT 4  
 OTX\_STRPU STANDARD; PRT; 371 AA.  
 AC Q26417;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Homeobox protein OTX (SPOTX).  
 GN OTX.  
 OS Strongylocentrotus purpuratus (Purple sea urchin).  
 CC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 CC Echinoidea; Euechinozoa; Echinacea; Echinoidae; Strongylocentrotidae;  
 CC Strongylocentrotus.  
 NCBI\_TaxID=7668;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95180495; PubMed=7875376;  
 RA Gan L., Mao C.-A., Wikramanayake A., Angerer L.M., Angerer R.C.,  
 RA Klein W.H.;  
 RT "An orthodenticle-related protein from Strongylocentrotus  
 RT purpuratus.";  
 RL Dev. Biol. 167:517-528(1995).  
 CC -1- FUNCTION: MAY PLAY A ROLE IN ACTIVATION OF THE SPEC2A GENE, BINDS  
 CC TO THE TATCC MOTIF WITH HIGH SPECIFICITY. MAY HAVE ADDITIONAL  
 CC FUNCTIONS IN THE DEVELOPING EMBRYOS.  
 CC -1- SUBUNIT: BINDS DNA AS A MONOMER.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DEVELOPMENTAL STAGE: FOUND INITIALLY IN ALL CELLS OF THE CLEAVING  
 CC EMBRYO, BUT GRADUALLY BECOMES RESTRICTED TO ORAL ECTODERM AND  
 CC ENDODERM CELLS.  
 CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.  
 CC "BICOID" SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: S76899; AAB33568.1; -;  
 CC HSSP: P06601; 1FJL.

DR TRANSFAC: T02416; -  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox.1.  
 DR PRINTS: PR00024; HOMEBOX.  
 DR ProDom: PD000010; Homeobox.1.  
 DR SMART: SM00389; HOX.1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 DR Homeobox: DNA-binding; Developmental protein; Nuclear protein.  
 KW DNA\_BIND  
 FT 131 190 HOMEBOX.  
 SQ SEQUENCE 371 AA; 41215 MW; B6BD493D684A33F7 CRC64;  
 Query Match 19.4%; Score 191.5; DB 1; Length 371;  
 Best Local Similarity 29.9%; Pred. No. 3.7e-07;  
 Matches 52; Conservative 29; Mismatches 64; Indels 29; Gaps 6;  
 QY 14 LSVYQYKISPTPOLGAASSAE-----GHVQGAPGIMGNMDEGVN-----55  
 DB 25 LNKPKYKLE---RLGMSSSPRLTIDCGDGRSP--VPSHMERPGGARVPMHLVPPY 79  
 QY 56 -HENGNNRGGMIPEGGGNQEPPOPPPEPPOAAMEGPOPEMNP---RTRTTKF 110  
 DB 80 AVSNPMYGEA-LPADRHLPPQGHMFPQVYLGPMTSERPHSNGIDPRKQRRERTTF 138  
 QY 111 TLLOVELESVFHRTQYPPVTRRELAEENLGVEDKVRWFKKRRACRKHORE 164  
 DB 139 TRAQLDVLLEFSTRYPRDIFMREEVAMKINLPESHVQWFKNRRAKCRQOOOO 192  
 RESULT 5  
 PTX3\_HUMAN  
 ID PTX3\_HUMAN STANDARD; PRT; 302 AA.  
 AC 07364;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Pituitary homeobox 3 (Homeobox protein PTX3).  
 GN PTX3 OR PTX3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBITaxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND VARIANT ADCC ASN-13.  
 RC TISSUE=Craniofacial;  
 RX MEDLINE=98282096; PubMed=9620774;  
 RA Semina E.V., Ferrell R.E., Mintz-Hittner H.A., Bitoun P.,  
 RA Alward W.L.M., Reiter R.S., Funkhauser C., Daack-Hirsch S.,  
 RA Murray J.C.;  
 RT "A novel homeobox gene PTX3 is mutated in families with autosomal-  
 dominant cataracts and ASD.";  
 RL Nat. Genet. 19:167-170(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Muscle;  
 RA Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: MAY PLAY A ROLE IN NORMAL ANTERIOR-CHAMBER AND LENS  
 CC DEVELOPMENT.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN DEVELOPING EYE LENS.  
 CC -1- DISEASE: MUTATIONS IN PTX3 APPEAR TO BE THE CAUSE OF THE ANTERIOR  
 CC SEGMENT MESENCHYMAL DYSGENESIS (ASMD) AND AUTOSOMAL-DOMINANT  
 CC INVOLVING THE FIRST (CORNEAL ENDOTHELIUM AND TRABECULAR MESHWORK),  
 CC SECOND (CORNEAL STROMA) AND THIRD (IRIS STROMA) MESENCHYMAL WAVES  
 CC OF NEURAL CREST. THE ASMD PHENOTYPE IS CHARACTERIZED BY CORNEAL  
 CC OPACITIES WITH OR WITHOUT IRIS ADHESIONS IN 100%, CATARACTS OF  
 CC VARYING SEVERITY IN 100% AND OPTIC-NERVE ABNORMALITIES IN 20% OF  
 CC AFFECTED INDIVIDUALS. ADCC IS CHARACTERIZED BY DOMINANT  
 CC TRANSMISSION OF A PHENOTYPE CONSISTING OF BILATERAL CONGENITAL  
 CC CATARACTS IN A MOTHER AND SON WITHOUT CLINICAL ANTERIOR-SEGMENT  
 CC ANOMALIES.

CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.  
 CC "BICOID" SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF041339; AAC24502.1; -  
 DR EMBL: BC011642; AAH11642.1; -  
 DR HSSP: P06601; 1FTJ.  
 DR TRANSFAC: T04311; -  
 DR GeneW: HGNC:9006; PTX3.  
 DR MIM: 602669; -  
 DR MIM: 107250; -  
 DR InterPro: IPR003654; Homeo\_OAR.  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox.1.  
 DR PRINTS: PR00024; HOMEBOX.  
 DR ProDom: PD000010; Homeobox.1.  
 DR SMART: SM00389; HOX.1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS50071; HOMEBOX\_2; 1.  
 DR PROSITE: PS50803; OAR.1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Disease mutation.  
 FT DNA\_BIND 62 121 HOMEBOX.  
 FT DOMAIN 262 275 OAR.  
 FT DOMAIN 240 250 POLY-ALA.  
 FT DOMAIN 268 272 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT VARIANT 13 13 S->N (IN ADCC).  
 FT /FTID=VAR\_003767.  
 SQ SEQUENCE 302 AA; 31832 MW; 1E5259206A8C2E87 CRC64;  
 Query Match 19.3%; Score 190.5; DB 1; Length 302;  
 Best Local Similarity 34.5%; Pred. No. 3.5e-07;  
 Matches 49; Conservative 23; Mismatches 45; Indels 25; Gaps 5;  
 QY 26 QLGASSASBEGHVGAGPLGMGNMPEGVNHENGMNMDGMIPEGGGNQEPPOPP 85  
 DB 2 EFLILSEAEAR---SPALSL---DAGTPHP-----QLPEHCKGGEHSD-----40  
 QY 86 EEPQAAMEGPOPEN-----MQPTRRKTKFTLLQVEELESVFRTQYPPVTRRELAEENLG 141  
 DB 41 SEKASASLPBGSPEDSLKKKORRORHTFTSQLOLELAEFTFQNNRYPDMSFREEIAVWTN 100  
 QY 142 VTEDKVRWFKKRRACRKHOR 163  
 DB 101 LTFARYRVWFKNRRAKRRKR 122  
 RESULT 6  
 ALX4\_MOUSE  
 ID ALX4\_MOUSE STANDARD; PRT; 399 AA.  
 AC 035137;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Homeobox protein aristal-less-like 4 (ALX-4).  
 GN ALX4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBITaxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Embryo;  
 RX MEDLINE=98086222; PubMed=9426253;



RA Qu S., Li L., Wisdom R.;  
 RT "Alx-4: cDNA cloning and characterization of a novel paired-type  
 RT homeodomain protein";  
 RL Gene 203:217-223(1997).  
 RN (2)  
 RN VARIANT LST GLN-206.  
 RX MEDLINE-98301426; PubMed-9636085;  
 RA Qu S., Tucker S.C., Ehrlich J.S., Levorse J.M., Flaherty L.A.,  
 RA Wisdom R., Vogt T.F.;  
 RT "Mutations in mouse Aristalless-like4 cause Strong's Luxoid  
 RT polydactyly";  
 RL Development 125:2711-2721(1998).  
 CC -1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN SKULL AND LIMB  
 CC DEVELOPMENT.  
 CC -1- SUBUNIT: BINDS DNA.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN OSTEOBLASTS, NOT EXPRESSED IN  
 CC BRAIN, HEART, INTESTINE, KIDNEY, LIVER, MUSCLE, SPLEEN AND TESTIS.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED FROM E8.25 AND CONFINED TO  
 CC MESENCHYMAL CELLS THROUGHOUT THE EMBRYO DEVELOPMENT. EXPRESSION IS  
 CC SEEN AT SEVERAL SITES INCLUDING CRANIOFACIAL REGION, FIRST  
 CC BRANCHIAL ARCH AND ANTERIOR ASPECT OF THE LIMB BUD.  
 CC -1- DISEASE: DEFECTS IN ALX ARE THE CAUSE OF STRONG'S LUXOID (LST)  
 CC PHENOTYPE. AT HETEROZYGOSITY LST IS CHARACTERIZED BY PREAXIAL  
 CC ABNORMALITIES OF THE HINDFEET AND, VERY RARELY, OF THE FOREFEET.  
 CC HOMOCYOTES SHOW PREAXIAL POLYDACTYLY OF ALL FOUR LIMBS,  
 CC REDUCTIONS AND DUPLICATIONS OF THE RADIUS, ABSENCE OF THE TIBIA,  
 CC CRANIOFACIAL DEFECTS, REDUCTION OF THE PUBIS, AND DORSAL ALOPECIA.  
 CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF001465; AAC39943.1; -;  
 DR HSSP: P06601; 1FJL.  
 DR TRANSFAC: T02967; -;  
 DR MGD: MGI:108359; Alx4.  
 DR InterPro: IPR003654; Homeo\_OAR.  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox; 1.  
 DR ProDom: PD000010; Homeobox; 1.  
 DR SMART: SM00389; HOX; 1.  
 DR PROSITE: PS00027; HOMEBOX\_1; 1.  
 DR PROSITE: PS00071; HOMEBOX\_2; 1.  
 DR PROSITE: PS50803; OAR; 1.  
 DR Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation; Activator; Disease mutation.  
 FT DNAS\_BIND 202 261  
 FT DNAS\_BIND 379 392  
 FT DOMAIN 206 206  
 FT VARIANT 206 206  
 FT R -> Q (IN LST; ABOLISHES DNA BINDING AND  
 FT TRANSCRIPTIONAL ACTIVATION).  
 SQ SEQUENCE 399 AA: 42762 MW: 28DE19DDACA21D25 CRC64:  
 Query Match 19.2%; Score 189.5; DB 1; Length 399;  
 Best Local Similarity 30.1%; Pred. No. 5.5e-07;  
 Matches 55; Conservative 22; Mismatches 49; Indels 57; Gaps 7;  
 QY 23 PTPQ---LGAASAEHGVOGAPGLMGNNPREGVNHENGMNDGMIPEGGGGNDPEPRQ 79  
 Db 95 PTPQPPAPAPPAHLYLORGA---CKTPPPDSLR-----LQSGSGHNAALQ 139  
 QY 80 QP-----OPPPEPA-----OAMEGPO-----PEMNO----- 102  
 Db 140 VPCYAKESNIGEPPELPDPSEPVGMDSYLSVKRTGAKGPDQDRAAELEPSELETDSESK 199  
 QY 103 --PRTTRKFTLLQVBELESVFRRTOYPDVPTRELAENLGVTEDKRVWVFKNKRARCR 160

Db 200 GKRRNRRTFTTSQLEBELEVFQKTHYPDYAREQJLAMPRTDLEAVQWQFQRRKMKR 259  
 QY 161 HQR 163  
 Db 260 RER 262  
 RESULT 7  
 ID PIX3\_MOUSE STANDARD; PRT; 302 AA.  
 AC 035160;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Pituitary homeobox 3 (Homeobox protein PITX3).  
 GN PITX3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Embryo, and Embryonic carcinoma;  
 RX MEDLINE-97472463; PubMed-9328475;  
 RA Semina E.V., Reiter R.S., Murray J.C.;  
 RT "Isolation of a new homeobox gene belonging to the Pitx/Rieg family:  
 RT expression during lens development and mapping to the aphakia region  
 RT on mouse chromosome 19.";  
 RL Hum. Mol. Genet. 6:2109-2116(1997).  
 RN [2]  
 RP DEVELOPMENTAL EXPRESSION.  
 RX MEDLINE-98282096; PubMed-9620774;  
 RA Semina E.V., Ferrell R.E., Mintz-Hitner H.A., Bitoun P.,  
 RA Alward W.L.M., Reiter R.S., Funkhauser C., Daack-Hirsch S.,  
 RA Murray J.C.;  
 RT "A novel homeobox gene PITX3 is mutated in families with autosomal-  
 RT dominant cataracts and ASDM.";  
 RL Nat. Genet. 19:167-170(1998).  
 CC -1- FUNCTION: Appears to be involved in normal eye anterior-chamber  
 CC and lens development.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Highly expressed in developing eye lens.  
 CC -1- DEVELOPMENTAL STAGE: First expressed in the eye in day E10  
 CC post-cotum embryos. Throughout eye development, expressed in the  
 CC lens placode and forming lens pit. From day E12, also detected in  
 CC the midbrain region, tongue, incisor primordia, condensing  
 CC mesenchyme around the sternum and vertebrae and in the head  
 CC muscles.  
 CC -1- DISEASE: Mutations in PITX3 appear to be the cause of the aphakia  
 CC phenotype, a recessive homozygous disease characterized by small  
 CC eyes and closed eyelids.  
 CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.  
 CC "BIOID" SUBFAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: AF005772; AAB87380.1; -;  
 DR HSSP: P06601; 1FJL.  
 DR TRANSFAC: T02666; -;  
 DR MGD: MGI:1100498; Pitx3.  
 DR InterPro: IPR003654; Homeo\_OAR.  
 DR InterPro: IPR001356; Homeobox.  
 DR Pfam: PF00046; homeobox; 1.  
 DR PRINTS: PR00024; HOMEBOX.

```
DR ProDom: PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
DR PROSITE; PS50803; OAR; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DNA_BIND 62 121
FT DOMAIN 262 275
FT DOMAIN 240 250
FT DOMAIN 268 272
SQ SEQUENCE 302 AA; 31714 MW; EB6EF6863B349264 CRC64;

Query Match
Best Local Similarity 40.6%; Pred. No. 5.3e-07;
Matches 41; Conservative 19; Mismatches 31; Indels 10; Gaps 2;

OY 67 IPGGGNGNEPPOPEPPEPQAAMEGPOPEN---MOPTRRTKFTLLQVEELSEVF 122
DB 28 LPEHCCKGGEHSD-----SEKASASLPGSGSPEDSLKKKQRQRTHTFSQQLQELATF 81

OY 123 RHTQYPDVPTRELAENLGVTEDKVRWFKNKRACRRQR 163
DB 82 QNRYPDMSTREIAVWNLTEARVWFKNRRARAKRRER 122

RESULT 8
PIX3_RAT
ID PIX3_RAT STANDARD; PRT; 302 AA.
AC P81062;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pituitary homeobox 3 (Homeobox protein Ptx3).
GN Ptx3 OR Ptx3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Hypothalamus;
RX MEDLINE=98058810; PubMed=9371841;
RA Smidt M.P., van Shaick H.S.A., Lantot C., Tremblay J.T., Cox J.J.,
  van der Kleij A.A.M., Wolterink G., Drouin J., Burbach J.P.H.;
  "A homeodomain gene Ptx3 has highly restricted brain expression in
  mesencephalic dopaminergic neurons.";
  Proc. Natl. Acad. Sci. U.S.A. 94:13305-13310(1997).
RL [2]
RN RP REVISIONS TO 66; 108; 122 AND 132-135.
RA Smidt M.P.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Appears to be involved in normal eye anterior-chamber
CC and lens development. Transcription factor which may be involved
CC in developmental determination of the mesod.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Restricted to mesencephalic dopaminergic
CC system (MESD).
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC "BICOID" SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ011005; CAA09455.2; -
CC HSSP; P06601; 1FTL.
CC ProDom: PD000010; Homeobox; 1.
CC TRANSFAC; T04305; -
```

```
DR InterPro; IPR003654; Homeo_OAR.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom: PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
DR PROSITE; PS50803; OAR; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DNA_BIND 62 121
FT DOMAIN 262 275
FT DOMAIN 240 250
FT DOMAIN 268 272
SQ SEQUENCE 302 AA; 31728 MW; DE0E81863C4AE714 CRC64;

Query Match
Best Local Similarity 40.6%; Pred. No. 5.3e-07;
Matches 41; Conservative 19; Mismatches 31; Indels 10; Gaps 2;

OY 67 IPGGGNGNEPPOPEPPEPQAAMEGPOPEN---MOPTRRTKFTLLQVEELSEVF 122
DB 28 LPEHCCKGGEHSD-----SEKASASLPGSGSPEDSLKKKQRQRTHTFSQQLQELATF 81

OY 123 RHTQYPDVPTRELAENLGVTEDKVRWFKNKRACRRQR 163
DB 82 QNRYPDMSTREIAVWNLTEARVWFKNRRARAKRRER 122

RESULT 9
CRX_BOVIN
ID CRX_BOVIN STANDARD; PRT; 299 AA.
AC Q9XSK0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cone-rod homeobox protein.
GN CRX.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Retina;
RA Zhu X., Craft C.M.;
  "Interaction of phosducin and Phlopi with CRX: potential
  transcriptional regulation function.";
  Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RL [2]
RN RP REVISIONS TO 66; 108; 122 AND 132-135.
RA Smidt M.P.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS AND TRANSCRIPTIVATES THE SEQUENCE 5'-TATC[CA]-3',
CC WHICH IS FOUND UPSTREAM OF SEVERAL PHOTORECEPTOR-SPECIFIC GENES,
CC INCLUDING THE OPSIN GENES. ESSENTIAL FOR THE MAINTENANCE OF
CC MAMMALIAN PHOTORECEPTORS.
CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).
CC -1- TISSUE SPECIFICITY: RETINA.
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF154123; AAD34645.1; -
CC HSSP; P06601; 1FTL.
CC TRANSFAC; T03489; -
CC InterPro; IPR001356; Homeobox;
CC Pfam; PF00046; homeobox; 1.
CC ProDom: PD000010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
```

```
DR PROSITE: PS00027; HOMEBOX_1; 1.
KM Transcription regulation; Activator; Homeobox; DNA-binding;
FT DNA_BIND 39 98 HOMEBOX.
SQ SEQUENCE 299 AA; 32253 MW; 512ED6A6DAFBAC19 CRC64;

Query Match
Best Local Similarity 41.5%; Score 187; DB 1; Length 299;
Matches 39; Conservative 16; Mismatches 27; Indels 12; Gaps 2;

OY 83 PPPEPAQA-AAEQPOPEMMP-----RTRRTKFTLLQVELESPFRHTQYPDV 130
DB 7 PGPHTSVNMLALSGPSVDLHRAVSPAPRKQRRRTFTTSQLEELALRAKTYQYPDV 66

OY 131 PPRRELAENLGVTEKRVKVRKRRRCRRHORE 164
DB 67 YAREVALKINLPESRVQYFKNRRAKRCQORQO 100

RESULT 10
GSC_HUMAN STANDARD; PRT; 252 AA.
ID GSC_HUMAN STANDARD; PRT; 252 AA.
AC P56915;
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein goosecoid.
GN GSC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leukocyte;
RX MEDLINE=94375063; PubMed=7916327;
RA Blum M., De Robertis E.M., Kojis T., Heinzmann C., Kilsak I.,
RA Gelisert D., Sparkes R.S.;
RT "Molecular cloning of the human homeobox gene goosecoid (GSC) and
RT mapping of the gene to human chromosome 14q32.1.";
RL Genomics 21:388-393(1994).
CC -1- FUNCTION: REGULATES CHORDIN (CHRD). MAY PLAY A ROLE IN SPATIAL
CC PROGRAMMING WITHIN DISCRETE EMBRYONIC FIELDS OR LINEAGE
CC COMPARTMENTS DURING ORGANOGENESIS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC "BICOID" SUBFAMILY.
DR HSSP: P06601; 1FTL.
DR TRANSFAC: T04037; -.
DR GeneW; HGNC:4612; GSC.
DR MIM: 138890; -.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; Homeobox.1.
DR ProDom: PD000010; Homeobox.1.
DR SMART: SM00389; HOX.1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
KM Developmental protein; Nuclear protein; DNA-binding; Homeobox.
FT DNA_BIND 155 214 HOMEBOX.
SQ SEQUENCE 252 AA; 27853 MW; 6C0EEC48C084D323 CRC64;

Query Match
Best Local Similarity 36.1%; Score 185.5; DB 1; Length 252;
Matches 44; Conservative 16; Mismatches 37; Indels 25; Gaps 2;

OY 65 GMPGGGGNPPRQPPPEPEPAQA-AAEQPOPEMMP----- 103
DB 96 GAVPPUGA--QQSCVPTPTGEGPSGVSVSPPHOMLYNVNGTISRTELQNLQHCRR 153

OY 104 --RTRRTKFTLLQVELESPFRHTQYPDVTPRRELAENLGVTEKRVKVRKRRRCRRH 161
DB 154 RRRRRRTITFDQALALENLFOETKYPRDVGTRQALARKVHLDEKYEYVFKRRKRRKRRQ 213
```

```
OY 162 OR 163
DB 214 KR 215

RESULT 11
GSC_MOUSE STANDARD; PRT; 256 AA.
ID GSC_MOUSE STANDARD; PRT; 256 AA.
AC Q02591;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein goosecoid.
GN GSC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92315328; PubMed=1352187;
RA Blum M., Gaunt S.J., Cho K.W.Y., Steinbeisser H., Blumberg B.,
RA Bittner D.A., de Robertis E.M.;
RT "Gastrulation in the mouse: the role of the homeobox gene goosecoid.";
RL Cell 63:1097-1106(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6;
RX MEDLINE=98079105; PubMed=94177125;
RA Danilov V., Blum M., Schweickert A., Campione M., Steinbeisser H.;
RT "Negative autoregulation of the organizer-specific homeobox gene
RT goosecoid.";
RL J. Biol. Chem. 273:627-635(1998).
CC -1- FUNCTION: GOOSECOID-EXPRESSING REGIONS OF THE GASTRULATING
CC MOUSE EGG CYLINDER HAVE ORGANIZER-LIKE ACTIVITY WHEN TRANSPLANTED
CC INTO XENOPUS EMBRYOS.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: IN EARLY GASTRULATION, EXPRESSED IN THE
CC DORSAL LIP. IN LATER STAGES OF DEVELOPMENT FOUND IN HEAD, LIMBS
CC AND BODY WALL.
CC -1- INDUCTION: BY ACTIVIN.
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC "BICOID" SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M85271; AAA37826.1; -.
DR EMBL: Y13149; CAAT73611.1; -.
DR EMBL: Y13150; CAAT73612.1; -.
DR PIR: A42768; A42768.
DR HSSP: P06601; 1FTL.
DR TRANSFAC: T02047; -.
DR MGD: MGI:95841; Gsc.
DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; Homeobox.1.
DR ProDom: PD000010; Homeobox.1.
DR SMART: SM00389; HOX.1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS50071; HOMEBOX_2; 1.
KM Developmental protein; Nuclear protein; DNA-binding; Homeobox.
FT DNA_BIND 160 219 HOMEBOX.
SQ SEQUENCE 256 AA; 27979 MW; 3639FB059AC3DB9E CRC64;

Query Match
Best Local Similarity 36.1%; Score 185.5; DB 1; Length 256;
Matches 44; Conservative 16; Mismatches 37; Indels 25; Gaps 2;
```



```

CC -1- FUNCTION: ACTS AS A TRANSCRIPTIONAL REGULATOR OF MUSCLE CREATINE
CC KINASE (MCK) AND SO HAS A ROLE IN THE ESTABLISHMENT OF DIVERSE
CC MESODERMAL MUSCLE TYPES. THE PROTEIN BINDS TO AN A/T-RICH ELEMENT
CC IN THE MUSCLE CREATINE ENHANCER.
CC
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; PMX1-A AND PMX1-B (SHOWN
CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN MESODERMALLY DERIVED
CC CELL TYPES. DURING EMBRYOGENESIS, HIGHEST LEVELS OF EXPRESSION
CC ARE FOUND IN THE MESENCHYME AND PRECARTILAGE ELEMENTS OF THE FACE
CC AND HIND LIMBS. IN THE ADULT, EXPRESSION IS RESTRICTED TO SKELETAL
CC MUSCLE, HEART AND UTERUS.
CC
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING CARDIOGENESIS.
CC
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEOBOX FAMILY.
CC
CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L06502; AAA39672.1; -
CC EMBL: U03873; AAC52139.1; -
CC EMBL: X59725; CAA42410.1; -
CC EMBL: S82911; AAB46839.1; -
CC HSSP: P06601; 1FTJL.
CC TRANSFAC: T02060; -
CC TRANSFAC: T02061; -
CC TRANSFAC: T02966; -
CC MGD: MGI:97712; Ptx1.
CC
CC InterPro: IPR003654; Homeo_OAR.
CC InterPro: IPR001356; Homeobox.
CC Pfam: PF00046; homeobox; 1.
CC ProDom: PD000010; Homeobox; 1.
CC SMART: SM00389; HOX; 1.
CC PROSITE: PS00027; HOMEBOX_1; 1.
CC PROSITE: PS50071; HOMEBOX_2; 1.
CC PROSITE: PS50803; OAR; 1.
CC
CC Homeobox: DNA-binding; Developmental protein; Nuclear protein;
CC Alternative splicing; Phosphorylation.
CC
CC FT DOMAIN 94 153 HOMEBOX.
CC FT MOD_RES 222 235 OAR.
CC FT VARSPPLIC 197 197 PHOSPHORYLATION (POTENTIAL).
CC FT 200 245 SAMATYSATCANNSPAQGINMANSTIRLAKESLORNO
CC VPTVN -> RSSSLPRCCHGILHNGF (IN ISOBFORM
CC PMX1-A).
CC
CC SO SEQUENCE 245 AA; 27269 MW; FF867FDD856E1115 CRC64;

```

```

DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein aristaless.
DE
GN AL.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=93138380; PubMed=8093690;
RA Schneitz K., Spielmann P., Noll M.;
RT "Molecular genetics of aristaless, a prd-type homeo box gene involved
RT in the morphogenesis of proximal and distal pattern elements in a
RT subset of appendages in Drosophila.";
RL Genes Dev. 7:114-129(1993).
CC
CC -1- FUNCTION: INVOLVED IN THE MORPHOGENESIS OF PROXIMAL AND DISTAL
CC PATTERN ELEMENTS IN A SUBSET OF APPENDAGES. APPEARS ALSO TO HAVE A
CC ROLE IN EARLY IMAGINAL DISK DEVELOPMENT. ARISTALESS MUTANTS
CC DISPLAY A REDUCTION IN SIZE OF THE ARISTA AND SCUTELLUM, A
CC REDUCTION OR COMPLETE ABSENCE OF THE TARSAL CLAWS, IRREGULARITIES
CC OF THE STERNOPLERUAL BRISTLES AND OF THE WING VEIN, AND A BENDING
CC OF THE WING BLADE.
CC
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC
CC -1- TISSUE SPECIFICITY: EXPRESSED DURING EMBRYONIC DEVELOPMENT, IN
CC DISTINCT EPIDERMAL PATTERNS IN HEAD, THORAX AND ABDOMEN AND IN AN
CC ENDODERMAL PATTERN IN THE ANTERIOR INTESTINAL TRACT.
CC
CC -1- DEVELOPMENTAL STAGE: FIRST EXPRESSED IN 4-8 HR-OLD EMBRYOS, PEAKS
CC IN 8-12 HR-OLD EMBRYOS AND CONTINUES THROUGH DEVELOPMENT UP TO
CC LATE LARVAL STAGE. IN THE HEAD REGION, DETECTED AT STAGE 10 IN THE
CC MAXILLARY AND LABIAL SEGMENT PRIMORDIA, AND IN LATER STAGES, IN
CC THE PROSPECTIVE ANTENNAL AND MANDIBULAR SEGMENT. IN THE EPIDERMIS,
CC EXPRESSION IS SEEN FROM STAGE 11 IN THORACIC AND ABDOMINAL LATERAL
CC PATCHES. EXPRESSION IN THE INTESTINAL TRACT BEGINS AT STAGE 13,
CC CONTINUES THROUGH STAGES 14 AND 15 IN THE ENDODERM OF THE ANTERIOR
CC MIDGUT AND AT STAGE 16, IS FOUND IN THE POSTERIOR END. EXPRESSION
CC IN THE IMAGINAL DISKS IS SEEN FROM LATE THIRD-INSTAR LARVAE IN
CC THE PROSPECTIVE THORAX, CLAW ORGAN, ANTENNA, SCUTELLUM AND WING
CC BLADE.
CC
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEOBOX FAMILY.
CC
CC -1- SIMILARITY: CONTAINS 1 HOMEOBOX DOMAIN.
CC
CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L08401; AAA28840.1; -
CC HSSP: P06601; 1FTJL.
CC
CC FlyBase: FBgn0000061; al.
CC InterPro: IPR000047; HTH_Repressr.
CC InterPro: IPR003654; Homeo_OAR.
CC InterPro: IPR001356; Homeobox.
CC Pfam: PF00046; homeobox; 1.
CC PRINTS: PR00024; HOMEBOX.
CC PRODOM: PD000010; Homeobox; 1.
CC SMART: SM00389; HOX; 1.
CC PROSITE: PS00027; HOMEBOX_1; 1.
CC PROSITE: PS50071; HOMEBOX_2; 1.
CC PROSITE: PS50803; OAR; 1.
CC
CC Homeobox: DNA-binding; Developmental protein; Nuclear protein.
CC
CC FT DNA_BIND 85 144 HOMEBOX.
CC FT DOMAIN 251 283 OAR.
CC FT 290 360 PRO-RICH.
CC GLN/PRO-RICH.

```

```

FT  DOMAIN 362 366 POLY-PRO.
SQ  SEQUENCE 408 AA; 43644 MW; 99F5C8CF787FE99E CRC64;
Query Match 18.6%; Score 183; DB 1; Length 408;
Best Local Similarity 32.7%; Pred. No. 1.6e-06;
Matches 48; Conservative 22; Mismatches 69; Indels 8; Gaps 4;

QY 18 QVKSIPFPLQGAASSAECHV-GQAGPGLGMNMEGVNHNENRMDGMIPREGGGNOE 76
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 6 EIKLEELFQGEAKLHPDAVLVDRAV--GSSASAGAAALVYMSVSGG-APSGASGASG 61
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 77 PQQPQPPPEPAQAAMGPPQENMQPRTKTKFTLLQVEELSEVFRTQYDPVTRREL 136
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 62 GTNSVSDSGNSDCEADYAPK--RKQRRYRTFTSFQLEELKAFSRTHYPDVTRREL 118
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 137 AENLGVTEDKRVVFKFKRRCRRHOR 163
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 119 AMKIGLTARLQVFNRRAKRKQEK 145
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
ALX4_HUMAN
ID ALX4_HUMAN STANDARD; PRT: 411 AA.
AC Q9H161; Q9HAY9; Q96JN7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein aristales-like 4.
GN ALX4 OR KIAA1788.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
NCBI_TaxID=9606;
RX MEDLINE=20489877; PubMed=11017806;
RA Wu Y.-Q., Badano J.L., McCaskill C., Vogel H., Potocki L.,
RA Shaffer L.G.;
RT "Haploinsufficiency of ALX4 as a potential cause of parietal foramina
RT in the 11p11.2 contiguous gene-deletion syndrome.";
RL Am. J. Hum. Genet. 67:1327-1332(2000).
RN [12]
RP SEQUENCE FROM N.A., VARIANT PEM2 PRO-272, AND VARIANT PRO-102.
RX MEDLINE=20558895; PubMed=11106354;
RA Wuyts W., Cleiren E., Homfray T., Rasore-Quartino A., Vanheeracker F.,
RA Van Hal W.;
RT "The ALX4 homeobox gene is mutated in patients with ossification
RT defects of the skull (foramina parietalia permagna, OMIM 168500).";
RL J. Med. Genet. 37:916-920(2000).
RN [13]
RP SEQUENCE FROM N.A., VARIANTS R-35 AND P-102, AND VARIANT PEM2 Q-218.
RX MEDLINE=20578741; PubMed=11137991;
RA Mavrouglanis I.A., Antonopoulou I., Baxova A., Kutilek S., Kim C.A.,
RA Sugawara S.M., Salamanca A., Wall S.A., Morris-Kay G.M.;
RA Wilkie A.O.M.;
RT "Haploinsufficiency of the human homeobox gene ALX4 causes skull
RT ossification defects.";
RL Nat. Genet. 27:17-18(2001).
RN [14]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 8:85-95(2001).
CC -1- FUNCTION: TRANSCRIPTION FACTOR INVOLVED IN SKULL AND LIMB
CC DEVELOPMENT.
CC -1- SUBUNIT: BINDS DNA (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- TISSUE SPECIFICITY: EXPRESSION IS LIKELY TO BE RESTRICTED TO BONE.
CC FOUND IN PARIETAL BONE.

```

```

CC -1- DISEASE: DEFECTS IN ALX4 ARE THE CAUSE OF PARIETAL FORAMINA 2
CC (PEM2), ALSO KNOWN AS FORAMINA PARIETALIA PERMAGNA (FPP), A
CC DISEASE CHARACTERIZED BY OVAL DEFECTS OF THE PARIETAL BONES CAUSED
CC BY DEFICIENT OSSIFICATION AROUND THE PARIETAL NOTCH, WHICH IS
CC NORMALLY OBLITERATED DURING THE FIFTH FETAL MONTH. PEM2 IS ALSO A
CC CLINICAL FEATURE OF A CONTIGUOUS GENE-DELETION SYNDROME KNOWN AS
CC POTOCKI-SHAFER SYNDROME.
CC -1- SIMILARITY: BELONGS TO THE PAIRED HOMEBOX FAMILY.
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 OAR DOMAIN.
CC -1- CAUTION: REF.1 SEQUENCE CONTAINS A FRAMESHIFT DUE TO THREE
CC ERRONEOUS SINGLE-NUCLEOTIDE DELETIONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF294629; AAC33961.1; ALT_FRAME.
CC EMBL: AF308822; AAK38835.1;
CC EMBL: AF308823; AAK38835.1; JOINED.
CC EMBL: AF308824; AAK38835.1; JOINED.
CC EMBL: AF308825; AAK38835.1; JOINED.
CC EMBL: AF404888; CAC15060.1;
CC EMBL: AJ279074; CAC15120.1;
CC EMBL: AJ279075; CAC15120.1; JOINED.
CC EMBL: AJ279076; CAC15120.1; JOINED.
CC EMBL: AJ279077; CAC15120.1; JOINED.
CC EMBL: AB058691; BAB47417.1; ALT_INIT.
CC HSSP: P06601; IEFL.
CC TRANSFAC: T04012;
CC Genew: HGNC:450; ALX4.
CC MIM: 605420;
CC MIM: 168500;
CC MIM: 601224;
CC InterPro: IPR003654; Homeo_OAR.
CC InterPro: IPR001356; Homeobox.
CC Pfam: PF00046; homeobox; 2.
CC ProDom: PD000010; Homeobox; 1.
CC SMART: SM00389; HOX; 1.
CC PROSITE: PS00027; HOMEBOX_1; 1.
CC PROSITE: PS50071; HOMEBOX_2; 1.
CC PROSITE: PS50803; OAR; 1.
CC KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
CC Transcription regulation; Activator; Polymorphism; Disease mutation.
CC FT DN_BIND 214 273 HOMEBOX.
CC FT DOMAIN 391 404 OAR.
CC FT DOMAIN 104 115 POLY-GLN.
CC FT VARIAT 35 35 T->R.
CC FT VARIAT 102 102 S->P.
CC FT VARIAT 218 218 R->Q (IN PEM2).
CC FT VARIAT 272 272 R->P (IN PEM2).
CC FT VARIAT 272 272 R->P (IN PEM2).
CC FT CONFLICT 134 134 D->N (IN REF. 1).
CC FT SEQUENCE 411 AA; 44176 MW; 243CDDC23088AEDF CRC64;

Query Match 18.6%; Score 183; DB 1; Length 411;
Best Local Similarity 31.9%; Pred. No. 1.7e-06;
Matches 45; Conservative 22; Mismatches 34; Indels 40; Gaps 5;

QY 63 DQGM-IPGGGGGNQEPQ-----QPPPEPBA-----QAAMGPP-- 97
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 134 DGLKLGSSGSHSAALQVPCYAKSSLGPELPDSDTGVGDSSYLSVKRAGVGPQDR 193
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 98 -----PEMNO-----PRTKTKFTLLQVEELSEVFRTQYDPVTRRELAEVNG 142
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 194 ASSDLPFLERKADSSNGKKRRRTFTTSYQLEELKAVFKTHIPDYARQGLAMRDL 253
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Mon, Sep 8 '15:07:38 2003

**us-09-867-753-2.rsp**

Page 11

Qy	143	TEDKVRVWEKNNKRCRRHQ	163
		: : : : :  :	
Db	254	TEARVQVWFQNNRRAKWKRRER	274

Search completed: April 28, 2003, 21:14:05  
Job time : 24 secs

**This Page Blank (uspto)**



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2003, 16:24:32 ; Search time 54 Seconds

(without alignments)  
3276.898 Million cell updates/sec

Title: US-09-867-753-1

Perfect score: 577

Sequence: 1 tccaacatcagcgctccag.....tctacatcgtcgtgactag 577

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued\_Patents\_NA: \*  
1: /cgn2\_6/prodata/1/lna/5A\_COMB.seq: \*  
2: /cgn2\_6/prodata/1/lna/5B\_COMB.seq: \*  
3: /cgn2\_6/prodata/1/lna/6A\_COMB.seq: \*  
4: /cgn2\_6/prodata/1/lna/6B\_COMB.seq: \*  
5: /cgn2\_6/prodata/1/lna/PCITUS\_COMB.seq: \*  
6: /cgn2\_6/prodata/1/lna/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71	12.3	877	US-09-129-888-1	Sequence 1, Appl
2	56.8	9.8	4080	US-08-710-249-3	Sequence 3, Appl
3	56.8	9.8	4080	US-09-220-157A-3	Sequence 3, Appl
4	56.4	9.8	906	US-08-957-351-6	Sequence 6, Appl
5	56.4	9.8	1017	US-08-957-351-5	Sequence 5, Appl
6	55.4	9.6	2481	US-08-958-642-3	Sequence 3, Appl
7	55.4	9.6	2481	US-08-778-394-1	Sequence 1, Appl
8	55.4	9.6	2481	US-08-778-423A-3	Sequence 3, Appl
9	54.6	9.5	906	US-08-957-351-2	Sequence 2, Appl
10	54.6	9.5	1392	US-08-957-351-1	Sequence 1, Appl
11	52	9.0	1251	US-09-636-735A-1	Sequence 1, Appl
12	51.6	8.9	1275	US-08-958-642-1	Sequence 1, Appl
13	51.6	8.9	1275	US-08-778-394-3	Sequence 1, Appl
14	51.6	8.9	1275	US-08-778-423A-1	Sequence 1, Appl
15	48.2	8.4	838	US-07-590-894C-1	Sequence 1, Appl
16	48.2	8.4	7218	US-08-332-463-14	Sequence 14, Appl
17	47.8	8.3	410	US-09-319-648-45	Sequence 45, Appl
18	43.8	7.6	157	US-09-277-078-41	Sequence 41, Appl
19	43.4	7.5	2338	US-08-425-069-1	Sequence 1, Appl
20	43.4	7.5	2338	US-08-317-844B-1	Sequence 1, Appl
21	42.8	7.4	786	US-08-403-379A-2	Sequence 2, Appl
22	42.8	7.4	786	US-08-929-414-2	Sequence 2, Appl
23	42.4	7.3	1223	US-08-957-351-4	Sequence 4, Appl
24	42.4	7.3	1223	US-08-957-351-29	Sequence 29, Appl
25	42.4	7.3	1240	US-08-957-351-8	Sequence 8, Appl
26	41.8	7.2	1500	US-09-593-711A-10	Sequence 10, Appl
27	41.4	7.2	245	US-09-117-121-27	Sequence 27, Appl

28	41.2	7.1	467	US-08-712-948-6	Sequence 6, Appl
29	41	7.1	1021	US-09-095-117-5	Sequence 5, Appl
30	41	7.1	1030	US-09-095-117-7	Sequence 7, Appl
31	41	7.1	1995	US-08-425-069-3	Sequence 3, Appl
32	41	7.1	1995	US-08-317-844B-3	Sequence 3, Appl
33	41	7.1	4403765	US-09-103-840A-2	Sequence 2, Appl
34	40.2	7.0	985	5215895-1	Sequence 191, App
35	39.8	6.9	432	US-09-397-787-191	Sequence 1, Appl
36	39.8	6.9	15378	US-08-785-420-1	Sequence 1, Appl
37	39.4	6.8	409	US-09-319-648-22	Sequence 22, Appl
38	39.4	6.8	432	US-08-642-255-48	Sequence 48, Appl
39	39.4	6.8	756	US-08-642-255-50	Sequence 50, Appl
40	39.4	6.8	4403765	US-09-103-840A-2	Sequence 2, Appl
41	39.2	6.8	477	US-09-135-994-1	Sequence 1, Appl
42	39.2	6.8	4425	US-08-749-169A-1	Sequence 1, Appl
43	39.2	6.8	4425	US-09-130-032A-1	Sequence 1, Appl
44	38.8	6.7	292	US-09-117-121-29	Sequence 29, Appl
45	38.8	6.7	292	US-09-344-529-8	Sequence 8, Appl

## ALIGNMENTS

```
RESULT 1
US-09-129-888-1
: Sequence 1, Application US/09129888B
: Patent No. 6063912
: GENERAL INFORMATION:
: APPLICANT: CHUN, Jong Yoon
: TITLE OF INVENTION: Placenta trophoblast-specific gene
: FILE REFERENCE: 1942/29
: CURRENT APPLICATION NUMBER: US/09/129, 888B
: CURRENT FILING DATE: 1998-08-06
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: Wordperfect 6.1 Windows
: SEQ ID NO 1
: LENGTH: 877
: TYPE: DNA
: ORGANISM: mouse
US-09-129-888-1

Query Match      12.3%   Score 71;   DB 3;   Length 877;
Best Local Similarity 56.7%   Pred. No. 4.6e+09;
Matches 131;   Conservative 0;   Mismatches 100;   Indels 0;   Gaps 0;

QY 326 CAGCCAGCAACTCGCCGACGAGTTCACGCTGTTCAGTGGAGAGCTGGAAGTGT 385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 488 CAACTGCGGTACACAGCGACGAGTTCACGCTGTTCAGTGCATGACCTGAGCGCTT 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 386 TTCGACACACTCAATACCTGATGTGCCACAGAAGGAAGTTCGGAAGTAACTAGT 445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 548 TTCGAAGAGTCTCCTACCCAGCTTCGAGCAGAAGAGGATCTTGACATGATGGGT 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 446 GTGACTGAAGCAAGTCCGGTGTGTTAAAGATTAAGGCCACATGTAGCGCAT 505
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 608 GTGATGAATGTATGTGTCAGAAATTTGTTGATGTAGAGAGAGCCCTTCCACAGAAG 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 506 CAGAGAAATTAATGCTCGCAATGACTAGCTGACCAAGGAGAGCTGAGT 556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 668 AGGAGATGCTGATGTTCTGTGGAAGTCCGCTCTTCCCAAGAGAGCTCT 718
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
US-08-710-249-3
: Sequence 3, Application US/08710249
: Patent No. 5858777
: GENERAL INFORMATION:
: APPLICANT: Villeponteau, Bryant
: APPLICANT: Feng, Junli
: APPLICANT: Andrews, William H.
: APPLICANT: Adams, Robert R.
: TITLE OF INVENTION: Methods and Reagents for Regulating
```





```

? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/958,642
? FILING DATE:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/778,423
? FILING DATE: December 31, 1996
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2481 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: other nucleic acid
? DESCRIPTION: /desc = "Oligonucleotide"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 163..1470
? US-08-958-642-3

```

Query Match	9.68;	Score 55.4;	DB 2;	Length 2481;
Best Local Similarity	-55.48;	Pred. No. 4.7e-05;		
Matches 107; Conservative	0;	Mismatches 86;	Indels 0;	Gaps 0

[illegible]

RESULT 7  
 US-08-778-394-1  
 ; Sequence 1, Application US/08778394  
 ; Patent No. 6028184  
 ;  
 ; GENERAL INFORMATION:  
 ;  
 ; APPLICANT:  
 ;  
 ; TITLE OF INVENTION: NOVEL METHOD FOR TESTING THE  
 ; DIFFERENTIATION STATUS IN PANCREATIC CELLS OF A MAMMAL  
 ;  
 ; NUMBER OF SEQUENCES: 4

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (Epo)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/778,394
FILING DATE:
CLASSIFICATION: 514
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2481 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "oligonucleotide"
FEATURE:
NAME/KEY: CDS
LOCATION: 163..1470
US-08-778-394-1

```

Query Match	9.68;	Score 55.4;	DB 3;	Length 2481;
Best Local Similarity	55.48;	Pred. No. 4.7e-05;		
Matches 107; Conservative	0;	Mismatches 86;	Indels 0;	Gaps 0

QY	319	GAATATGACACCGAAACTGGGGCAGAAAGTTCAACGCTGTGGTCCAGGTGAGAGACTCTGA	378
Db	822	GAAACCGAAGCTCGAAGAATAGAAACATCTTTTACCCAAAGCAGCATTTGAGGCTCTGGA	881
QY	379	AAGTGTTCCTCGACACACTCAATACCTGTGTGCCAAGAAGGAAGCACTTGCCGAAA	438
Db	882	GAAAGAGTTTGAGAGGACCCATTATTCAGATGTGTTGGCCGGGAAAGACTAGCAGCCAA	941
QY	439	CTTAGGTGTACTGAAAGACAAAGTGGCGGTTTGCTTTAAAGATTAAGAGGCGCAGATGTAG	498
Db	942	AATGATCTCACTCGAAGCAAGAAATACAGATGATGTTTCTAATCGAAGGCGCCAAATGTGG	1001
QY	499	GGCAGATCAGAGA	511
Db	1002	AAGAGAGAGAAA	1014

```

1      RESULT 8
2      US-08-778-423A-3
3      : Sequence 3, Application US/08778423A
4      : Patent No. 6071697
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT:
9      :
10     : APPLICANT:
11     :
12     : TITLE OF INVENTION: NOVEL METHOD FOR TESTING THE
13     :
14     : TITLE OF INVENTION: DIFFERENTIATION STATUS IN PANCREATIC CELLS OF A MAMMAL
15     :
16     : NUMBER OF SEQUENCES: 16
17     :
18     : COMPUTER READABLE FORM:
19     :
20     : MEDIUM TYPE: Floppy disk
21     :
22     : COMPUTER: IBM PC compatible
23     :
24     : OPERATING SYSTEM: PC-DOS/MS-DOS
25     :
26     : SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
27     :
28     : CURRENT APPLICATION DATA:
29     :
30     : APPLICATION NUMBER: US/08/778,423A
31     :
32     : FILING DATE: December 31, 1996
33     :
34     : INFORMATION FOR SEQ ID NO: 3:
35     :
36     : SEQUENCE CHARACTERISTICS:
37     :
38     : LENGTH: 2481 base pairs
39     :
40     : TYPE: nucleic acid
41     :
42     : STRANDEDNESS: single
43     :
44     : TOPOLOGY: linear
45     :
46     : MOLECULE TYPE: other nucleic acid
47     :
48     : DESCRIPTION: /desc = "oligonucleotide"
49     :
50     : FEATURE:
51     :
52     : NAME/KEY: CDS
53     :
54     : LOCATION: 163..1470
55     :
56     : US-08-778-423A-3

```

[illegible]





NUMBER OF SEQUENCES: 16  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/778,423A  
FILING DATE: December 31, 1996  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1275 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "Oligonucleotide"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 166..1161  
US-08-778-423A-1

Query Match 8.9%; Score 51.6; DB 3; Length 1275;  
Best Local Similarity 52.9%; Pred. No. 0.00035;  
Matches 111; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 341 CGCACGAGTTACAGCTGTTGCGAGGTGAGAGCTGGAAAGTGTTCACACACTCAA 400  
DB 685 CGGCGCTCTCTCTCCCGGAGCAAGCCGAGGCACTGGAGAAAGATTTCAGCCGTGGGCGAG 744  
QY 401 TACCCCTGATGCGCCACAGAAAGGAGACTTGCCGAAACTTAGGTGAGTCAAGCAAA 460  
DB 745 TATCCAGATTCAGTGGCGCCCTGGGAAGCTGCTGCTCCACCTCTCTGCGTGAAGACAG 804  
QY 461 GTGCGGGTTGTTTAAAGTAATAAGGCGCAGATGTAGGCGACATCAGAGAAATTATG 520  
DB 805 GTGAGGGTTGGTTTCTTACAGAAAGAGCCAAATGGCGCAGGCAAGAGCTGAATG 864  
QY 521 CTGCGCAATGACTACTGCTGACCCAGAC 550  
DB 865 GAAGCACAGCTGCCAGGTGCTTCCAGGAC 894

RESULT 15  
US-07-590-894C-1  
Sequence 1, Application US/07590894C  
Patent No. 5342761  
GENERAL INFORMATION:  
APPLICANT: Macleod, Carol L  
TITLE OF INVENTION: A No. 5342761el Onco-Fetal Gene, Gene Product  
TITLE OF INVENTION: and Uses Therefor  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dr. Benjamin A. Adler  
STREET: 1301 McKinney St. 41st Floor  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77010-3095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/590,894C  
FILING DATE: 01-Oct-1990  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Adler Ph.D., Benjamin A.  
REGISTRATION NUMBER: 35,423  
REFERENCE/DOCKET NUMBER: D5237  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 713-651-5587  
TELEFAX: 713-651-5246  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 838 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
STRAIN: AKR1 Jackson  
INDIVIDUAL ISOLATE: SL12 cell line  
DEVELOPMENTAL STAGE: Bone marrow-adult  
TISSUE TYPE: Lymphoma  
CELL TYPE: T-cell  
CELL LINE: SL12.4 clone  
FEATURE:  
OTHER INFORMATION: N at nucleotide 838 represents a polyA string  
US-07-590-894C-1

Query Match 8.4%; Score 48.2; DB 1; Length 838;  
Best Local Similarity 49.8%; Pred. No. 0.0023;  
Matches 150; Conservative 0; Mismatches 148; Indels 3; Gaps 1;

QY 256 GCACGAGCCGAGCCCGCCGAGAGCCGCCCAAGCGCCCATGTGAGGCTCCGACGCC 315  
DB 366 GGAGCAGAGAACAAATGAGCCAGTTGCTGAGGGCAGTGAAGCCAGAGGAATGGAATCC 425  
QY 316 CGAAGACATGAGCCAGCAGAACTCGGCGACAGATTCACGCTGTGACAGTGGAGAGACT 375  
DB 426 TGGGGTAGAGAGATGCCCCCTCCAGGGCTGAGTTGCCCAAGCATGAGTGAAGGAACT 485  
QY 376 GGAAGTGTTCGACACACTCAATACCTGATGTGCCCAAGAAAGGAATTCGCCA 435  
DB 486 GAGTCCATTTTGCAGCGCACTAATTCCTTGATGT---CCCAAGGAGAGATCTTGATAG 542  
QY 436 AAAGTTGAGTGTGACTGAAGCAAGAGCGGCTTGTGAAGTAATAAGGCGCAGATG 495  
DB 543 ACTGATGATGCCGTGTGTCCAGAGTGCAGAAATGGTTAAGATCAGAGAGGCTGGCGC 602  
QY 496 TAGCGGACATCAGAGAAATTAATGCTGCCAATGACTACGCTGACCCAGACGACTG 555  
DB 603 CAGAAAGAAACAGAGAGAGGCGCAACACAGTCCCTGAAACATTTAGAGGAACATTGAGTG 662  
QY 556 T 556  
DB 663 T 663

Search completed: April 28, 2003, 19:25:28  
Job time : 72 secs

**This Page Blank (uspto)**